Investigation of multicomponent adsorption isotherms in chromatography using high-throughput formats

A thesis submitted for the degree of Doctor of Philosophy

by

Nicholas John Field

The Advanced Centre for Biochemical Engineering, Department of Biochemical Engineering, University College London

2017

I, Nicholas Field confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

_____

Abstract

Adsorption isotherms in chromatography are critical in determining the separation of solutes during column separations. Multicomponent protein adsorption isotherms, which are relevant during the downstream processing of biopharmaceutical products, have received limited study historically. The studies and methodologies which have been assessed have mainly focused on small, simple, chromophore containing proteins which have limited applicability to industrially relevant bio-therapeutics. The reasons why this area of study has received limited attention include the experimental effort associated with generating such large data sets as well as the difficulty in obtaining data of good enough quality.

The work explored here presents and optimises the deployment of high-throughput chromatography formats as well as automated liquid handling systems in order to elucidate adsorption isotherms of proteins. Additionally, alternative rapid analytical methods involving the collection of protein UV spectra in conjunction with multivariate data analysis have been applied to quantify protein mixtures. These rapid high-throughput methods decrease the experimental effort associated with multicomponent isotherm studies. 3 binary isotherms and 1 ternary isotherm have been studied for larger, non-chromophore containing model proteins. The propagation of error in single component and multicomponent isotherms has been investigated to understand what drives the propensity for error as well as methods to mitigate problematic regions of investigation.

The fitting of the multicomponent ion exchange isotherms across multiple salt levels to isotherm formalisms proved elusive which precluded their application for *in silico* modelling of column separation. Short of that a heuristic optimisation of a binary mixture was achieved quantifying eluted fractions using the UV spectra multivariate method.

Impact statement

Improving understanding and techniques associated with the elucidation of multicomponent protein adsorption isotherms in chromatography facilitates industrial partners and other academic groups performing research in this area. Historically, the area has received limited study, the work presented here decreases the experimental effort required to perform such studies and presents alternative methods of quantification to either corroborate data output from such studies or find less error prone ways to perform such studies.

Lowering the experimental effort for such studies could allow industrial partners to understand their manufacturing processes better, thus reducing risk and maximising profitability. These techniques also facilitate one of the major data collection processes associated with *in silico* chromatography separations and potentially eases its application to the optimisation of industrial manufacturing processes, potentially reducing the cost of process development and the time to market thus maximising profitability and expediting the availability of new drugs.

Acknowledgements

Firstly I would like to thank my academic supervisor Prof. Ajoy Velayudhan for his direction and support who has taught me many valuable lessons, applicable not only to the work discussed here, but also generally in how to manage my future professional career more generally. I'd also like to thank Dr. Spyridon Konstantinidis for his tireless patience, support and advice, especially in the early years of this journey. Additionally, I'd like to thank Kosma Jurlewicz for his camaraderie and terrible jokes throughout.

I gratefully acknowledge the financial support of the Engineering and Physical Sciences Research Council (EPSRC) and Centre of Doctoral Training for their support and teaching throughout this process.

I would like to thank my parents for giving me the start in life which has allowed me to get to this point. Although they were not able to guide me during the work undertaken here, the lessons they gave me earlier in life were critical in allowing me to reach this point. Although my mother, Patricia cannot be here to share in my progress I know she would be proud to see where her lessons have taken me.

Most importantly I would like to thank my wife Zakiyyah Ismail for her support and encouragement along every step of this process. There is absolutely no doubt that I would not have made it to this point were it not for her unwavering love and support.

1.1 Table of contents

Contents

1.2 List of figures

1.3 List of tables

1.4 Nomenclature

Abbreviation or symbol	Definition
A	Empirical constant in FLJ isotherm
AEX	Anion exchange
ALHS	Automated liqudi handling
ANOVA	One-way analysis of variants
API	Active pharmaceutical ingrediant
AU	Absorbance units
BSA	Bovine serum albumin
C_{Cl-}	Chloride ion concentration (mM)
CEC	Capillary electrochromatography
CEX	Cation exchange
Con	Conalbumin
C_0	Corrected starting concentration (mg/mL)
C_{start}	Starting liquid concentration (mg/mL)
C_{eq}	Liquid solute concentration at equilibrium (mg/mL)
C_{eq}*	Equilibrium liquid concentration from elution data (mg/mL)
C_i	Concentration in elution cycle I (mg/mL)
CV	Column volumes
Cyt-C	Cytochrome-C
DBC	Dynamic binding capacity
DE	Differential evolution algorithm
DiTi	Disposable tips
EP	Elution on a plateau (frontal analysis method)
F	Phase ratio (liquid/solid)
FLJ	Freundlich-Langmuir-Jovanovic
Hb	Haemoglobin
HIC	Hydrophobic interaction chromatography
HT	High-throughput
K	equilibrium constant
k'	Retention factor
LV	Latent variable
Lys	Lysozyme
MAD	Median absolute deviation
MCI	Multicomponent isotherm
m_{req}	Mass retained from equilibration stage (ug)
$m_{e,i}$	Mass in elution step *i* (ug)
$m_{re,i}$	Mass retained from elution step (ug)
$m_{r,i}$	mass retained from equilibrium/elution/wash step *i* (ug)
$m_{rw,i}$	Mass retained from wash step *i* (ug)
$m_{w,i}$	Mass in wash step *i* (ug)
Ova	Ovalbumin
PCA	Principle component analysis
pI	Isoelectric point
PLS	Partial lease squares regression or projection of latent structures
q	Adsorbed at equilibrium calculated using mass balance (mg/mL resin)

q^*	Adsorbed concentration at equilibrium calculated from elution (mg/mL resin)
q_0	concentration of charged ligands on the stationary phase in SMA
RMSEcv	Root mean squared error cross validation
RP	Reverse phase
RSD	Relative standard deviation
SCI	Single component isotherm
SEC	Size exclusion chromatography
SMA	Steric mass action
UHQ	Ultra High Quality (water)
VR	Volume range
V_L	Volume of liquid loaded for batch adsorption experiment (uL)
V_r	Volume retained between cycles (uL)
V_{resin}	Volume of resin (uL)
z	protein effective charge is the SMA model
α	Exponential term in FLJ isotherm
Λ	Maximum protein binding capacity
σ	Standard deviation
σ_S	Steric hindrance in the SMA model
phase	Phase ratio (solid/liquid)

Chapter 2 Introduction

2.1 Purification of Biopharmaceuticals

Protein therapeutics are of great and increasing importance in the treatment of disease. Within the top 30 best-selling protein therapeutics there are treatments for cancer, immune disease, diabetes, multiple sclerosis and haemophilia. Some of these therapeutics have been revolutionary, such as the use of insulin in the treatment of diabetes, epoetin in anemia and rituximab for non-Hodgkin lymphoma. Protein therapeutics are also of significant financial importance with 2010 sales in the USA and EU of $108bn (Dimitrov, 2012) and the value of this market is projected to grow, particularly in the case of monoclonal antibodies (Ecker, Jones, & Levine, 2015).

The earliest full scale biopharmaceutical processes were established 50 years ago and were reliant on precipitation using organic solvents in the production of blood plasma products such as polyclonal immunoglobulin G for the treatment of immunodeficiency diseases (Buchacher & Iberer, 2006). Generally speaking modern protein therapeutics are expressed in genetically modified bacterial, yeast, insect or mammalian cells, the active pharmaceutical ingredient (API) is then recovered from the cell culture broth before purification.

As therapeutic proteins are usually injected into the patient rather than through oral ingestion they must be of exceptionally high quality. This process is further confounded by the presence of critical impurities which can illicit powerful immunogenic responses or side effects in small quantities. Examples of such impurities include: endotoxin found in the cell walls of gram negative bacteria such as *E. coli* which are commonly used for target protein expression, host cell proteins which can be present even in processes using extracellularly expressed API due to cell lysis during cell culture, growth factors used during cell culture, DNA and RNA as well as aggregates which can not only illicit an immune response but also reduce shelf life by seeding aggregation. Some of these critical impurities such as aggregates can have physical properties very similar to the target compound. Chromatography offers a

powerful tool in the purification of biopharmaceuticals from process and product related impurities as it offers a high efficiency of separation. Moreover, modern chromatography resins, particularly ion exchange resins, can offer extremely high interaction strength and saturation capacity. This can be useful when purifying a product from a high volume low concentration feed, but also offers effective use of factory floor space as smaller volumes of solid phase can be used. This is especially true in comparison to liquid-liquid extractions which can require similar volumes of each phase (Carta & Jungbauer, 2010). An early reference to ion exchange is made in The Bible when Moses throws a tree into "bitter" water. It's been suggested that the water was contaminated with various salts and that the cellulose in the wood adsorbed these ions rendering the water potable (Streat, 1995).

2.2 Fundamentals solute-sorbent interaction at equilibrium and the effect mass transfer resistance

Adsorption isotherms represent interaction between solute and sorbent at equilibrium, single component isotherms are expressed in plots where the x-axis represents the liquid concentration and the y-axis the adsorbed concentration at equilibrium as shown in Figure 2-1a, b and c. Isotherms give information about the strength of interaction as well as the binding capacity of the sorbent. Interactions between protein solutes and chromatography sorbents used in their purification typically exhibit convex isotherm shapes represented in Figure 2-1 (1b). In the case of single component isotherms, the isotherm can be divided into two parts, the linear portion which is also equivalent to the Henry constant and the non-linear portion which explores the curvature and saturation capacity of the sorbent as sites on the sorbent are filled up and the solute is in competition with itself to bind (Carta & Jungbauer, 2010).

During analytical chromatography the aim of the separation is only to get information about the purity or mass of one or more of the components in the separation. As such analytical separations are carried out on the linear portion of the isotherm where elution behaviour is unaffected by factors such as load or the

presence of other species, thus simplifying any separation. Separation in this linear region is represented in Figure 2-1 (1a), the greater slope of component B in this system results in later elution in Figure 2-1 (2a). Figure 2-1 (3a) shows what the resulting peak would look like after the kinetics of mass transfer and non-idealities of flow have caused the solute plug to spread. In an ideal separation, solutes travelling through a column under flow could be imagined as a plug with sharp front and rear boundaries. In reality resistances to mass transfer and mobile phase dispersion effects disperse the front and back of the peak giving peaks a Gaussian appearance as shown in Figure 2-2 (Carta & Jungbauer, 2010; Schmidt-Traub, 2005).

The mass transfer effects are caused by the time taken for solute to travel through the chromatography bead in order to bind or elute. The significant forms of mass transfer resistance are: The external mass transfer resistance created by a layer of mobile phase which is stationary and not subject to flow which the solute must travel through via diffusion, the solute must also travel through the porous space in the bead via pore diffusion, there can also be kinetic resistance to binding which is usually ignored as it is often comparatively negligible, and finally diffusion in the adsorbed state. In addition to this axial diffusion in the interstitial space and hydrodynamic dispersion also contribute to band spreading (Carta & Jungbauer, 2010; Schmidt-Traub, 2005).

During preparative chromatography the goal is to recover a component from the separation either attempting to maximise the mass and/or purity of the component, as such preparative chromatography is often performed in non-linear regions of the isotherm. In such regions the shape of the isotherm affects peak shape. As curvature is seen in the isotherm the slope flattens with increasing liquid concentration, which is associated with decreased retention, as such increasing solute concentration within the column is associated with decreased slope and retention. This results in an elution band with a sharp front and diffuse tail represented in Figure 2-1 (2b). When mass transfer resistance and non-idealities of flow are taken into account the peak resembles what is seen in Figure 2-1 (3b). Concave isotherms seen in Figure 2-1 (1c) are sometimes observed during gas chromatography but are less common during

liquid chromatography, particularly in the case of protein separations (Conder & Young, 1979; Schmidt-Traub, 2005).

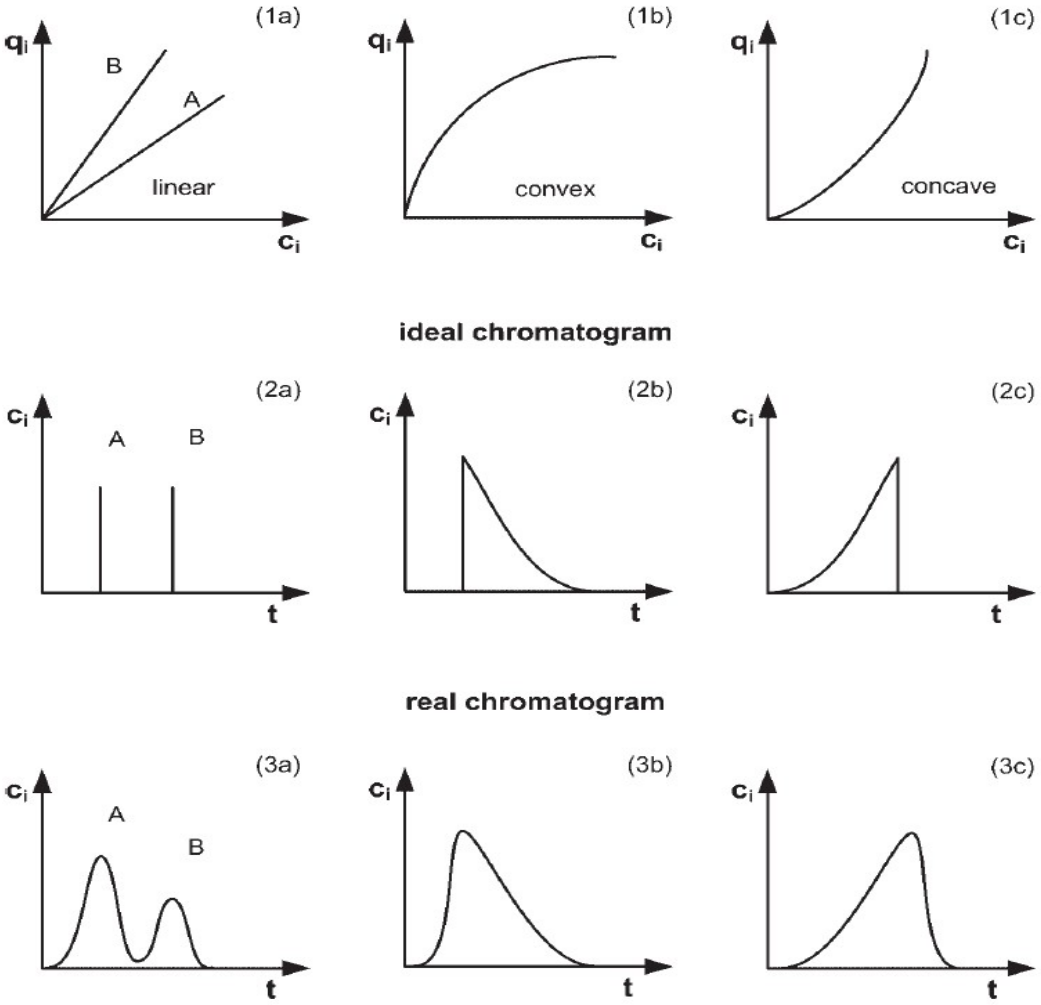

Figure 2-1 Pictures showing the influence of isotherm shape upon peak shape. Panels 1a, 1b and 1c present different isotherm shapes, panels 2a, 2b and 2c present the peak shape those isotherms produce in an ideal situation with no band spreading, 3a, 3b and 3c represent the sample peaks after some band spreading has occurred. Reproduced from (Schmidt-Traub, 2005).

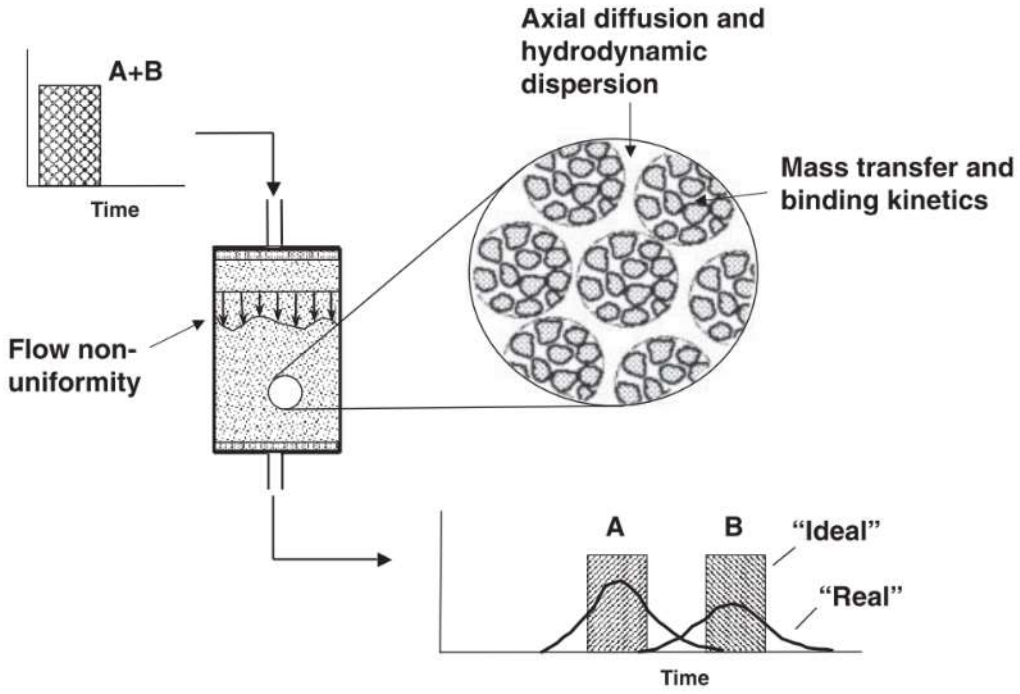

Figure 2-2 Effect of mass transfer resistances on peak shape. Reproduced from (Carta & Jungbauer, 2010).

In order to model a separation an isotherm description is required in order to link mobile and stationary phase concentration during separation. The most common formalisms used in solid-liquid separations involving proteins are the Langmuir, Langmuir-Freundlich and steric mass action (SMA) models. An isotherm description does not necessarily require any physical basis behind its structure and it can be arbitrary, it merely needs to be able to capture the trends but simple enough to accurately estimate the parameters (Guiochon, Felinger, & Shirazi, 2006; Guiochon & Katti, 1987). However, the addition of parameters which lead to unobserved or unlikely behaviour should be avoided.

The SMA model is a popular choice of isotherm description for ion exchange separations as the salt concentration counter-ion is a parameter in the equation, meaning the equation can be deployed directly at different NaCl concentrations, see equation 4-1. Additionally the parameters fit in the single component case can theoretically be applied to multicomponent cases eliminating the need to generate time consuming multicomponent adsorption isotherms. In reality the complex

multifaceted interactions at play in protein interactions mean this approach is often unsuccessful and multicomponent data is required to fit the parameters.

It is also possible to link the Langmuir and Freundlich-Langmuir isotherms to the concentration of the salt via an exponential term and an extra parameter for the salt concentration, this isotherm will be referred to as the Freundlich-Langmuir-Jovanovic (FLJ) isotherm (J.-X. Huang & Horváth, 1987), the Freundlich-Langmuir equation reduces to Langmuir when α=1. Across the isotherm equations 4-1 and 4-2 q is the adsorbed protein concentration at equilibrium and C_{eq} is the liquid protein concentration at equilibrium and C_{cl^-} is the counter ion concentration, in this case it's for an anion exchanger so chloride ions have been included, in the case of a cation exchanger sodium ions might be substituted. In the SMA formalism q_0 is the concentration of charged ligands on the stationary phase, z is the protein effective charge, σ_s is the steric hindrance and in the case of the adapted FLJ isotherm Λ represents the maximum protein adsorption capacity, K is the equilibrium constant, α and A are empirical constants.

$$q = \frac{K.(q_0 - (z + \sigma_s).q)^z C_{eq}}{(C_{Cl^-})^z} \qquad 2\text{-}1$$

$$q = \frac{\Lambda.K.{C_{eq}}^{\alpha}.\exp(-\mathrm{A}.C_{Cl^-})}{1 + K.{C_{eq}}^{\alpha}.\exp(-\mathrm{A}.C_{Cl^-})} \qquad 2\text{-}2$$

Figure 2-3 shows how the isotherms change in shape both in different Cl^- conditions and as the coefficients change for the SMA isotherm, Figure 2-4 shows an equivalent pair of plots for the FLJ isotherm. It is clear that across both isotherms increasing chloride ions decreases the binding capacity of the isotherm and also flattens the slope of the linear portion of the isotherm and that changing the values of the parameters allows a range of isotherm saturation capacities and shapes to be explored.

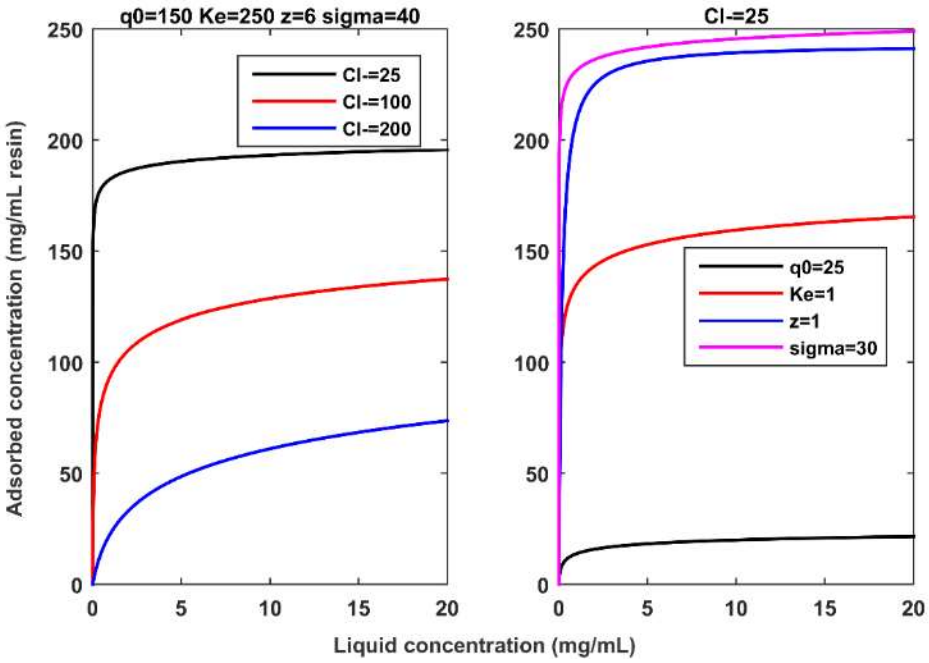

Figure 2-3 Examples SMA isotherm. Left hand panel shows the isotherm with the same parameters at different Cl^- concentrations. The parameters are displayed in the title and different levels of Cl^- displayed in the legend. Right hand panels display isotherms at the same Cl^- concentration as shown in the title and different parameters which are displayed in the legend. The base parameters for the right hand panel are the same as what is displayed in the title of the left had panel with one parameter changed as displayed in the legend.

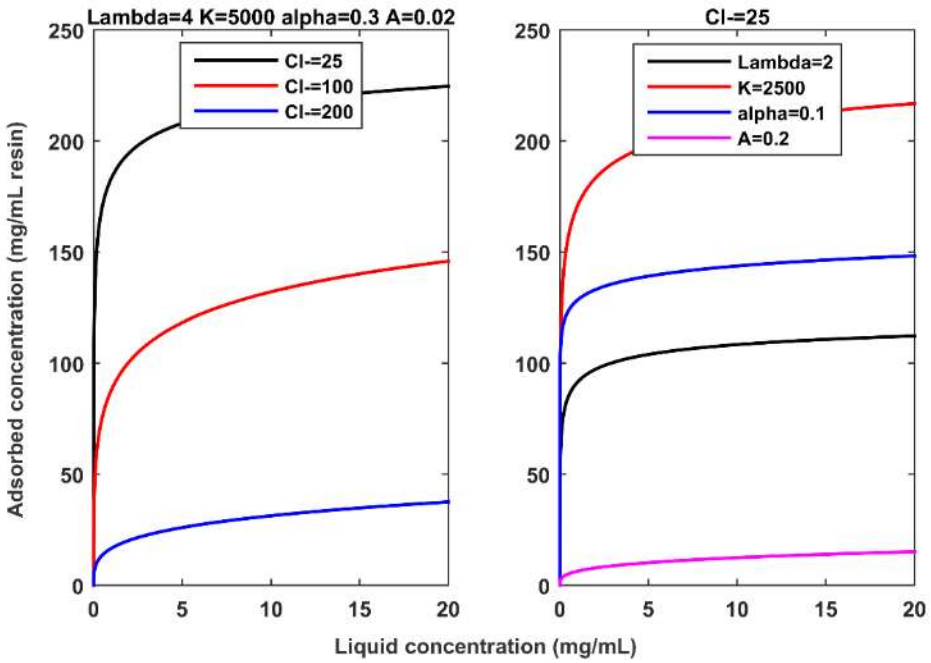

Figure 2-4 Example Langmuir-Freundlich-Jovanovic isotherms. Left hand panel shows the isotherm with the same parameters at different Cl^- concentrations. The parameters are displayed in the title and different levels of Cl^- displayed in the legend. Right hand panels display isotherms at the same Cl^- concentration as shown in the title and different parameters which are displayed in the legend. The base parameters for the right hand panel are the same as what is displayed in the title of the left had panel with one parameter changed and displayed in the legend.

2.3 Multicomponent isotherms

The useful function of chromatography is to take a mixture of compounds and to separate them either in order to analyse their contents or recover component(s) at increased purity or concentration from the starting mixture. Single component isotherms are useful in modelling when looking at separation in a non-overloading state i.e. during analytical separation or a low loading preparative run. But if the desire is to maximise throughput of a column step it is desirable to run the column approaching its saturation capacity. During downstream processing of biopharmaceuticals the feed is always a combination of components, although in late stage separations the impurities may be product related and very similar to the target compound. This being the case the relevant isotherms will be multicomponent, not single component.

In multicomponent isotherms multiple solutes are competing for binding space on the sorbent, solutes which interact more strongly will displace less strongly retained species, less strongly interacting solutes may also be able to displace more retained species if their concentration is high enough. Multicomponent isotherm descriptions for the SMA model are displayed in equation 2-3 and equation 2-4 displays the multicomponent case for the adapted Langmuir-Freundlich isotherm for component *i* in a solution containing *j* number of components.

$$q_i = \frac{K_i.\left(q_0 - \sum_{j=1}^{M}\left(z_j + \sigma_{S,j}\right).q_j\right)^{z_i}}{(C_{Cl^-})^{z_i}} C_{eq,i} \qquad 2\text{-}3$$

$$q_i = \frac{\Lambda_i.K_i.{C_{eq,i}}^{\alpha_i}.\exp(-\mathrm{A}_i.C_{Cl^-})}{1 + \sum_{j=1}^{N} K_j.{C_{eq,j}}^{\alpha_j}.\exp(-\mathrm{A}_j.C_{Cl^-})} \qquad 2\text{-}4$$

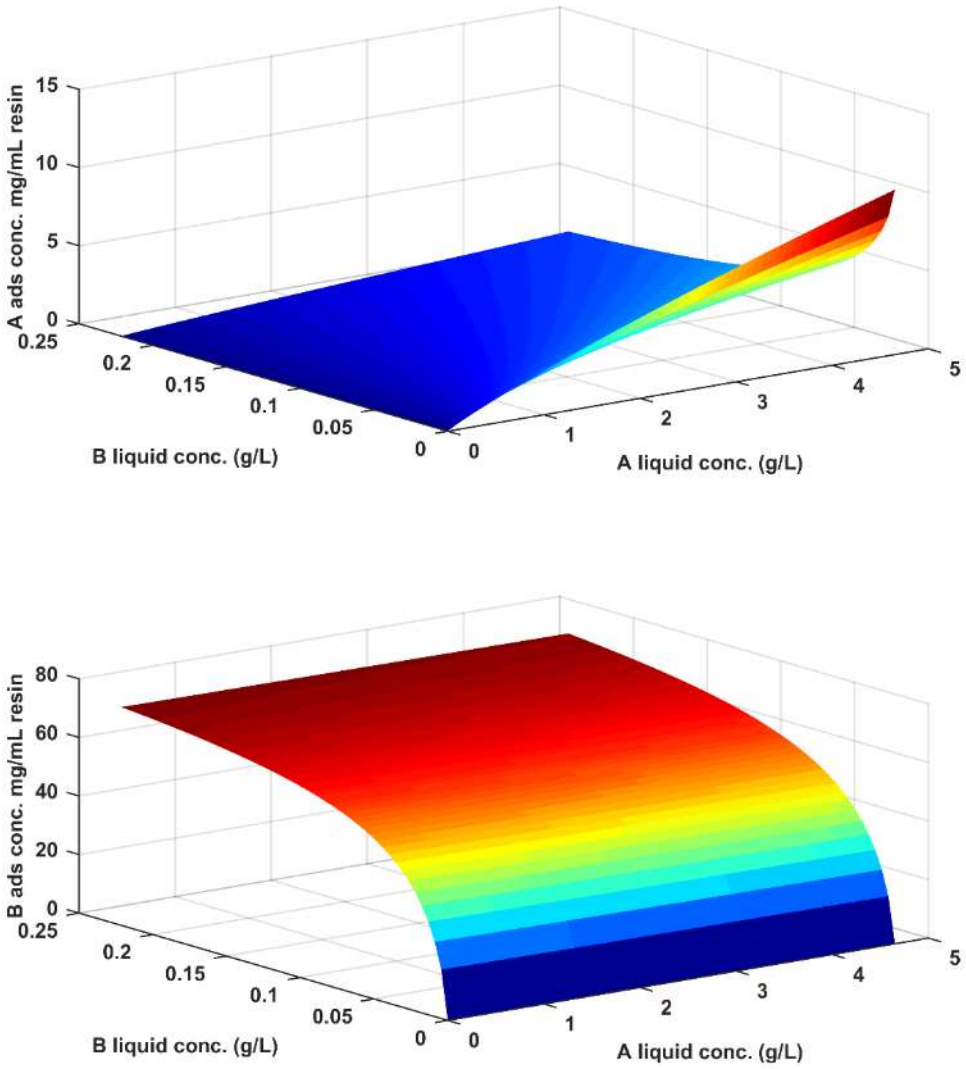

Figure 2-5 Example binary isotherms displaying adapted Langmuir-Freundlich isotherms. Parameters are as follows: Λ_1=2, K_1=2, α_1=0.9, A_1=0.03, Λ_2=2, K_2=1000, α_2=0.8, A_2=0.04, Cl^-=0.

Figure 2-5 displays an example binary (2 component) isotherm for components A and B. There are two halves of the isotherm; one displaying the adsorbed concentration for each component. In this case the *K* value has been set much higher for component B than A in order to show how one component displaces the other. The top panel shows how, with increasing liquid concentration of component A, the adsorbed concentration of A also increases, which is what was also observed for the single component isotherms. Additionally as the concentration of B increases A is displaced from the resin. In the bottom panel displaying adsorbed concentration of B the adsorbed concentration increases with increasing liquid concentration of B. In this case A is very poor at displacing B from the resin and nearly no effect is seen. Multicomponent isotherms can of course contain any number of components.

2.4 Different isotherm study approaches and their utility

There are two common methods for carrying out isotherm studies; frontal analysis and batch adsorption experiments. Frontal analysis is carried out on columns under flow and it involves providing a steady feed of solute to the column which migrates down the column with a sharp front (as represented in Figure 2-1 (2a)). This front is detected and can be analysed in order to ascertain isotherm properties. Frontal analysis can also be performed when the feed is cut which produces a boundary where the solute concentration is returning to 0, in fact a chromatogram peak can be thought of as a pair of frontal boundaries back to back with the first increasing and the second decreasing. In both cases the boundaries travel more slowly than the liquid phase and the ratio between the velocity between the liquid phase and boundary gives information about the partition between the liquid and solid phases, which is the slope in the isotherm. By running experiments at different loads both the linear and non-linear parts of the isotherm can be explored. Frontal analysis can be performed on multicomponent feeds where each component will present a step in the boundary, which is dependent on the partitioning of the component. As described in Figure 2-1 (2b) the shape of the boundary is dependent upon the curvature of the isotherm when at non-linearity. Ideally, a frontal analysis experiment done in the linear region should present a perfectly sharp front. However, mass transfer resistance and axial diffusion will cause that front to soften (Figure 2-1 (3a)), this can complicate finding the isotherm parameters in the overloading case (Conder & Young, 1979). An example result from a binary frontal analysis experiment is displayed in Figure 2-6.

The alternative method of isotherm study is the batch adsorption method where known quantities of solute and solid phase are introduced to the liquid phase in a closed system where there is no flow. The solute is allowed to reach equilibrium and partition between the liquid and solid phases before the liquid phase is assayed for any remaining solute in order to ascertain the adsorbed concentration via a mass balance calculation. Multiple solute loads are required in order to explore the linear portion as well as curvature and the saturation capacity. Unlike the frontal analysis

method no further mathematical transformation is required to elucidate the isotherm as the directly assayed concentration represents the C_i and is plotted in the x-axis, the adsorbed value q_i on the y-axis. Mass transfer does not complicate studies so long as the isotherm has been equilibrated properly.

2.4.1 Multicomponent frontal analysis in isotherm determination

Binary isotherms of p-resol and phenol on a reverse phase sorbent have been studied in research lead by Horvath (Jacobson, Frenz, & Horvath, 1987). The two methods of studying multicomponent isotherms using frontal analysis previous to this work were either by performing frontal analysis using a radiolabelled isotope to identify the multiple fronts or via the elution on a plateau (EP) method. In the EP method a constant feed of multicomponent solute is applied creating a plateau, a perturbation of that feed is then applied and the associated elution peak is then monitored (Conder & Young, 1979). Both approaches have their limitations: isolation of a suitable isotope can be difficult or costly, in the case of studying proteins labelling a protein with a suitable identifying tag which does not affect its chromatographic behaviour would also present a challenge. Moreover the EP methods requires equipment of uncommonly broad response range and sensitivity. The alternative methods presented by Jacobson *et al.* are 2 fold; firstly multicomponent isotherms were elucidated by performing frontal analysis on a binary feed and regularly sampling the resulting elution plateaus via analytical HPLC, an example binary frontal analysis result is displayed in Figure 2-6. The second method required no such analytical HPLC but instead assumes Langmuirian behaviour and used the elution volume of the elution steps in order to parameterise the Langmuir isotherm description. Whilst both approaches are appealing they do present difficulties especially in the proposed context of multicomponent protein adsorption isotherms in ion exchange chromatography.

The two proposed methods require many frontal analysis column runs, many in overloading conditions, 37 were performed in total. The high value of biopharmaceutical products would make a study of this nature prohibitively

expensive in addition to being rather time consuming in an environment where process development times are squeezed in order to maximise time on the market before patent expiry (Bensch, Schulze Wierling, von Lieres, & Hubbuch, 2005). The assumption of Langmuirian behaviour would be unwise in the case of protein adsorption isotherms as it is well known their behaviour at equilibrium can be rather complex and difficult to capture even when using more complex formalisms (Roth, Unger, & Lenhoff, 1996). A more recent publication in the same area investigating simple aromatic and aliphatic ternary mixtures points out that in order to successfully study multicomponent isotherm via frontal analysis very efficient columns are required with minimal band spreading (Lisec, Hugo, & Seidel-Morgenstern, 2001). Ideal separations with negligible band spreading can be challenging in the context of proteins as their increased size and decreased diffusivity can mean band spreading and mass transfer resistance are often not negligible. Taken together it would seem frontal analysis is a tool not best suited to the study of multicomponent protein adsorption isotherms.

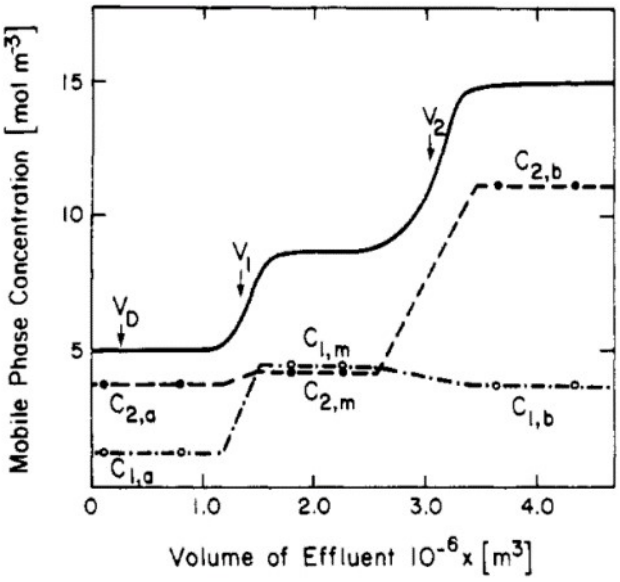

Figure 2-6 Diagram showing an example result from a binary frontal analysis experiment. Solid line represents the result generated by the optical detector. Dashed lines represent concentration of component 1 (C_1) and component 2 (C_2) at the beginning of the experiment when the feed is equilibrated (a), in between the two fronts (m) and after both fronts (b). V_D is the dead volume, V_1 and V_2 are the two fronts.

2.4.2 Batch adsorption and high-throughput chromatography formats

As discussed previously the alternative method of studying isotherms is to perform a batch adsorption experiment. In this format known quantities of solute and solid phase are introduced to the liquid phase in a closed system where there is no flow, the solute is allowed to reach equilibrium and partition between the liquid and solid phases before the liquid phase is assayed for any remaining solute in order to ascertain the adsorbed concentration via a mass balance calculation. An attractive feature of this format is the apparent simplicity in that the assayed concentration at equilibrium represents the x-axis of the isotherm and the calculated adsorbed concentration represents the y-axis, and so long as the isotherm is left to incubate for sufficiently long enough the mass transfer kinetics do not obfuscate the results in any way. The drawbacks of this format include the requirement for many experimental conditions to be run at different combinations of multicomponent loads (this increases with the number of components under study) in order to explore the linear range, as well as curvature in addition to any displacement within an isotherm. The number of experimental conditions also increases as different liquid phase conditions often need to be tested. In the case of an ion exchanger not only do different levels of mobile phase modifier need to be explored (normally different NaCl concentrations) but different pH levels may also be tested. Additionally this format of experimentation can be rather error prone with significant errors introduced when attempting to aliquot sorbent as well as complications arising from absolute quantification of solutes in a mixture which can be very challenging for proteins when baseline HPLC separation is not possible (Di Marco & Bombi, 2001; Seidel-Morgenstern, 2004).

Traditionally batch adsorption experiments involved manually aliquoting equilibrated slurry suspensions of resin into sealed reaction vessels along with known quantities of solute in liquid phase, this can be done at µL or mL scale (Fritz & Schluender, 1974; G L Skidmore & Chase, 1990; Graham L. Skidmore, Hortsmann, & Chase, 1990). As previously mentioned this manual activity can increase the random and systematic error associated with such experiments. In order to maximise data

quality it is advisable to use large resin volumes. However, that is at odds with the desire to use small quantities of valuable solutes such as in the case of proteins.

More recent innovations have led to the creation of slurry plates. Slurry plates are 96 well plates which contain a small volume of chromatography resin in every well. There are different types of this plate available depending on the required use/budget. Plates can be operated in a dead end format where a conventional deep square well plate has chromatography resin dispensed into it. The liquid and solid phases are separated simply by gravity settling followed by aspiration of the liquid phase after incubation. Alternatively filter plates can be used which allow separation of solid and liquid phases either via vacuum filtration or the use of a centrifuge. Slurry plate can either be self-dispensed, often with the aid of an automated liquid handling system or they can be purchased pre-dispensed such as in the case of PreDictor™ plates supplied by GE Healthcare™ (Chhatre & Titchener-Hooker, 2009). Self-dispensed plates can also be achieved with aid of a vacuum pump with specialised equipment which leaves a predetermined volume of adsorbent plaque in each well (Herrmann, Schroder, & Hubbuch, 2006). The increased lead time associated with optimising the reliable dispensing of resin aliquots is traded off against the decreased costs in consumables over time and the increased flexibility meaning the desired volume of any resin available can be aliquoted rather than only what resin/volume combinations are commercially available. This flexibility has allowed the creation of 384 well plates for even further increases in throughput (Kittelmann, Ottens, & Hubbuch, 2015). Slurry plates offer advantages not just in their potential increased throughput but also in the improved accuracy and precision of resin aliquotation in comparison to manual handling.

Automated liquid handling systems (ALHS) are often used in conjunction with slurry plates (Nfor et al., 2010). An ALHS is essentially a robotic pipette able to transfer predetermined volumes of liquid from one location to another, it is often equipped with a robotic arm enabling it to transfer labware, such as 96 well plates. Finally it can also be equipped with additional modules such as a plate mixer or plate reader, an annotated picture of the ALHS used in this work is displayed in Figure 4-1.

The use of slurry plates on automated liquid handling systems (Coffman, Kramarczyk, & Kelley, 2008; Traylor, Xu, Li, Jin, & Li, 2014; Welsh et al., 2014) brings additional benefits such as: Increased throughput, the potential for walkaway automation and improved repeatability (Gu & Deng, 2007; Rhode et al., 2004; Xie, Wang, Carpenter, & Wu, 2004). Slurry plates deployed in conjunction with ALHS are therefore well suited to the determination of single and multicomponent adsorption isotherms.

2.4.3 The application of slurry plates for multicomponent isotherm study

Self-dispensed slurry plates have been used in the generation of multicomponent isotherms of sugars before *in silico* modelling of their separation was executed (Chilamkurthi, Willemsen, van der Wielen, Poiesz, & Ottens, 2012). Single component, binary and ternary isotherms were generated only exploring the linear region of the isotherm for D-glucose, D-galactose, L-arabinose, lactose and sugar acid on four different ionic forms of sorbent: Ca^{2+}, K^{+}, Na^{+} and H^{+}. The isotherms were screened in order to assess if there was any competitive or co-operative (anit-Langmuirian see Figure 2-1 (1c)) behaviour in this linear region before the *K* values of the single component isotherms were assessed in order to identify the resins with the greatest selectivity of binary and ternary mixtures. Once a suitable resin had been identified column experiments were performed for the ternary mixture and the regressed binary isotherm parameters accurately reflected the ternary separation.

This work illustrates the powerful potential of HT chromatography formats in conjunction with ALHS for the study of multicomponent isotherms. However, the limitation of experiments to the linear portion of the isotherm significantly eases the challenging nature of multi-component isotherm study and means this work has limited applicability to protein adsorption studies. The limitation in scope of the above work to linear isotherms not only eases the challenge of fitting complex multicomponent, multidimensional isotherm shapes to descriptions but it also means the data is less error prone. The plateau of the isotherm is most susceptible to random error and can be difficult to capture in more weakly interacting isotherms as

will be discussed in Chapter 6, and is also touched upon elsewhere (a. Osberghaus et al., 2012).

The power of automated liquid handling with slurry plates was illustrated by another group with respect to proteins (Staby et al., 2007). The increased throughput offered by the format was leveraged in order to test 6 weak anion exchange resins using 5 proteins across different buffer conditions. Physical characteristics of the resins such as particle size distribution as well as chromatographic behaviour such as static binding capacity were screened. Slurry plates and ALHS were successfully deployed in order to parameterise the SMA isotherm from BSA and liploase single component isotherm data. This was successfully used in conjunction with pulse experiments to capture non-idealities and optimise a binary separation of BSA and lipolase (Staby et al., 2007). Whilst this work certainly illustrates the potential of HT formats much of the work was done at larger scale and only a smaller subset using slurry plates. Additionally the authors were fortunate enough to have a system where the single component parameters of the SMA model could be applied to the binary mixture for a successful optimisation. This allowed them to avoid having to generate multicomponent isotherm generation which presents many additional challenges around analytics, quantification and catastrophic experimental error (Di Marco & Bombi, 2001; a. Osberghaus et al., 2012). Successful application of SMA parameters from the single component case to the multicomponent case cannot always be assumed as proteins are large complex solutes with complex interaction with the sorbent and with one another (Golshan-Shirazi & Guiochon, 1994). In such cases, which are likely to be the norm for large complex biopharmaceuticals, multicomponent data is required.

2.4.4 Current state of multicomponent protein adsorption isotherms

A review of multicomponent protein adsorption isotherms reveals common problems and trends across different publications and groups. Generally the approach taken to parameterise multicomponent isotherm descriptions is to fit parameters using single component isotherms before generating a subset of multicomponent data in order to validate the model. This is to avoid the extremely laborious task of generating multicomponent data in multiple buffer conditions. There are examples of this approach being successful (Faraji, Zhang, & Ray, 2015; Fargues, Bailly, & Grevillot, 1998; Rege, Tugcu, & Cramer, 2003; Tao, Carta, Ferreira, & Robbins, 2011; Zhou, Su, & Sun, 2007), but this approach is often either unsuccessful or does not capture the isotherm behaviour fully (Aboudzadeh, Aboudzadeh, Jiawen, & Bin, 2007; Finette, Baharin, Mao, & Hearn, 1997; Lewus & Carta, 1999; Liang, Fieg, Shi, & Sun, 2012; G L Skidmore & Chase, 1990; Xu & Lenhoff, 2009; S. Zhang & Sun, 2003). It has been suggested that the reason why this approach fails is due to the complex protein-protein interactions that occur during multi component binding which are not captured during single component experiments (Xu & Lenhoff, 2009). The proteins studied in many of these publications are often model proteins of limited size and complexity such as: lysozyme (14 kDa) cytochrome-C (12 kDa), ribonuclease-a (14 kDa) and cytochrome-b5 (15 kDa) (Baumann, Huuk, Hahn, Osberghaus, & Hubbuch, 2016; Cano, Offringa, & Willson, 2005; Demin, Mogilevskaya, & Samsonov, 1997; Garje, Hartmann, Papamicmichael, Deckwer, & Anspach, 1999; Rege et al., 2003; G L Skidmore & Chase, 1990; Xu & Lenhoff, 2009). As such one might expect that these complicating protein-protein interactions to become even more confounding in the biopharmaceutical arena where the most common products are immunoglobulins which have a molecular weight of approximately 150 kDa. Until the complex underlying nature of these interactions is better understood to allow better extrapolation into the multicomponent space it seems chromatographers are at least partially reliant on multicomponent isotherm generation.

The laborious, time consuming and meticulous nature of multicomponent isotherm generation is repeatedly referenced across the literature and the difficulty of generating data of sufficient quality is also touched upon (Xu & Lenhoff, 2009). There have been limited attempts to understand the source and underlying nature of the errors associated with multicomponent isotherm study despite the fact that the errors associated with single component isotherm study have been reasonably well studied (a. Osberghaus et al., 2012). Additionally, binary adsorption isotherms have received some attention (LeVan & Vermeulen, 1981; Lewus & Carta, 1999; Xu & Lenhoff, 2009), but a very small number of publications have attempted to generate ternary adsorption isotherms (Close, Salm, Bracewell, & Sorensen, 2014; Melter, Ströhlein, Butté, & Morbidelli, 2007; Tao et al., 2011).

Slurry plates represent an excellent format for isotherm study as they alleviate many of the difficulties of the batch adsorption method in this type of study: They are parallelised and high throughput in format meaning they alleviate the difficulties associated with many experimental loads/conditions. Resin is usually aliquoted in an automated fashion meaning issues around repeatability are reduced. Assuming the process of placing the resin in aliquots has been optimised properly accuracy of this format is also improved in comparison to preparing resin by hand in larger quantities for each experiment. The suitability of slurry plates for isotherm study has been exploited (Chilamkurthi et al., 2012; Nfor et al., 2010) but their application to multicomponent isotherms has received less attention (Baumann et al., 2016).

Therefore, there exists an opportunity to exploit the high-throughput nature of slurry plates in conjunction with ALHS in order to study not only binary and ternary isotherms but also gain a better understanding of the underlying factors that make their study so challenging. The proteins chosen for this study are bovine serum albumin (BSA), ovalbumin (Ova) and conalbumin (Con) and the chromatography sorbent is Capto^{TM} Q which is a strong anion exchanger. This combination of sorbent and proteins was chosen because they represent larger proteins which have typically received limited study in this area, they are also highly competitive on anion

exchangers with overlapping elution peaks and they can be produced at gram scale for reasonable cost.

An underlying requirement for multicomponent isotherm study is an analytical method which is able to assay individual protein concentrations in a mixture which will be explored in the section below.

2.5 HPLC separation of model mixture

The traditional method for analysis of individual protein titres from a mixture is to ideally achieve baseline separation of all the proteins that require quantification using analytical HPLC runs, proteins are then quantified based on the peak area in reference to a standard curve. Baseline separation is an important caveat because if there is overlap between the peaks the assayed concentration of one protein becomes partially dependant on the concentration of another (Janson & Jönsson, 2011). During the analysis of proteins baseline separation is often not achieved. For example, in the quantification of low weight molecular species during analysis of active product during manufacture of IgG using size exclusion chromatography, or characterisation of acidic and basic species of IgG (Tao et al., 2011; Toueille, Uzel, Depoisier, & Gantier, 2011). For the application of an analytical method to multicomponent isotherms the assaying of concentrations independent of the presence of other proteins is important, without this the isotherm could become systematically over or underestimated in certain regions. If baseline separation cannot be achieved it is possible to apply deconvolution methods in order to quantify the contribution of each peak in the overlapping region and quantify it, although such approaches are not always successful (Di Marco & Bombi, 2001).

A second prerequisite for an effective analytical method is efficient recovery of the protein from the column. Irreversible interactions, or interactions which are not disrupted during a single typical analytical gradient, can lead to underestimation in the sample during the analytical run and the appearance of ghost peaks in subsequent analytical runs as previously retained protein elutes (Daly, Gilar, Gebler,

Corporation, & R, 2003). The feasibility of the 5 most common analytical HPLC separation methods will be investigated for the purpose of quantification of BSA-Ova-Con mixtures: reverse-phase HPLC (RP-HPLC), hydrophobic-interaction HPLC (HIC-HPLC), size exclusion chromatography HPLC (SEC-HPLC), cation-exchange HPLC (CEX-HPLC) and anion exchange HPLC (AEX-HPLC).

2.5.1 Analytical separation using reverse phase HPLC and hydrophobic interaction chromatography

RP-HPLC relies on hydrophobic interactions between solute and sorbent. Solute binds to the stationary phase using a polar liquid phase such as water. The mobile phase is gradually changed, typically in a gradient, by increasing the non-polar composition and thus decreasing the polar content. Typically the mobile phase modifier is acetonitrile although ethanol and isopropanol are commonly used. This causes the proteins to elute, less hydrophobic proteins elute first followed by more hydrophobic proteins (Pettersson, 2011).

The interaction between solute and stationary phase during HIC has been described as “an interaction of molecules with each other which is stronger than the interaction of the separate molecules with water and which cannot be accounted for by covalent, electrostatic, hydrogen bond or charge-transfer forces” (Gutsche, 1970). Alternatively it can be characterised as the strong interaction between water molecules excluding the interaction between water molecules and the solute (Hjertén, Yao, Eriksson, & Johansson, 1986). In practice retention of solute is achieved by binding it to the stationary phase in a liquid phase which is high in salt content, salts which are higher in the Hofmeister series tend to be more effective. The solute is then eluted by decreasing the salt content of the mobile phase. Again less hydrophobic proteins elute first followed by more hydrophobic proteins

Investigation into reverse phase separation of BSA, ovalbumin and conalbumin suggested that whilst baseline separation at least between Ova and BSA would be possible, as shown in Figure 2-7 there are issues with recovery of Ova. These recovery issues lead to the underestimation of Ova during the analytical run and the

appearance of ghost peaks during subsequent analytical runs (Daly et al., 2003). Whilst it may have been possible to improve recovery by using isopropanol instead of acetonitrile as mobile phase B, separation of Con from BSA and Ova would also have been challenging (Carr, 2002; Fausnaugh, Kennedy, & Regnier, 1984). Alternatively investigation of separation via HIC suggested that separation of Ova and Con by this method would be also be challenging as retention times were within 0.2 min as shown in Table 2-1 (Fausnaugh et al., 1984).

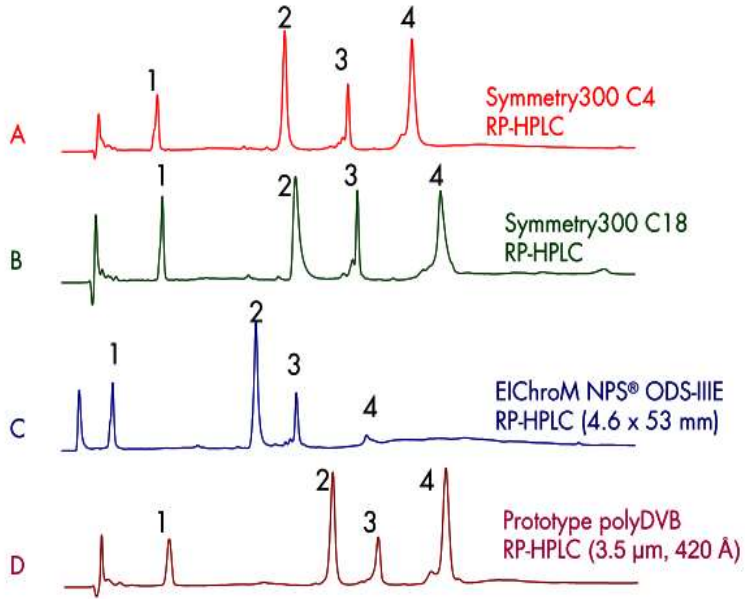

Figure 2-7 Reverse phase separation of model proteins on different Waters columns. 1) RNAse 2) BSA 3) β-lactoglobulin 4) Ova (Daly et al., 2003).

Protein	*Retention time (min)*
Cytochrome C	0.6
Myoglobin	0.8
RNAse	1.6
Con	6.3
Ova	6.5
Lysozyme	8.5
B-glucosidase	15.6
α-Chymotrypsin	16.6
α-Chymotrypsinogen-A	18.1
Lactoperoxidase	19.5
BSA	20.5
Ferritin	20.8

Table 2-1 Retention times of model proteins on HIC using the same buffer conditions and gradient. Table reproduced from (Fausnaugh et al., 1984).

2.5.2 Analytical separation using size exclusion chromatography

SEC-HPLC separates solutes based on their hydrodynamic volume which is mostly dependent upon the molecular weight of the species. The stationary phase consists of a cross linked polymer which creates porous beads. The solute is then able to access the porous space within the bead to varying degrees dependant on the size of the solute. As the mobile phase is pumped through the solid phase smaller molecules which can access more of the pore space are retarded in their passage through the column and elute after larger solutes which can access less of the pore space, or none of it at all (Hagel, 2011).

	BSA	*Ova*	*Con*
Molecular mass (kDa)	67	45	77
Isoelectric point (pH units)	4.8	4.5	6.6
*Price (£/g)**	3	18	154
Solubility (g/L)	> 40	> 40	> 40
Source	Bovine blood	Hen egg white	Hen egg white

*Table 2-2 Protein properties of BSA, Ova and Con. *Prices are approximate, subject to change and not necessarily the price paid in this work as bulk purchase discounts were available.*

The different molecular mass of the 3 proteins described in Table 2-2. This data suggest that SEC-HPLC may be a tractable method of analytical separation. Although baseline separation of BSA (66 kDa) and Ova (44 kDa) (Figure 2-8) was achieved using a UPLC with 1.7 µm bead size it is unlikely that baseline resolution between BSA (66 kDa) and Con (77 kDa) could be achieved (Waters Corporation, 2017). The difference in molecular mass between BSA and Con is 11 kDa as opposed

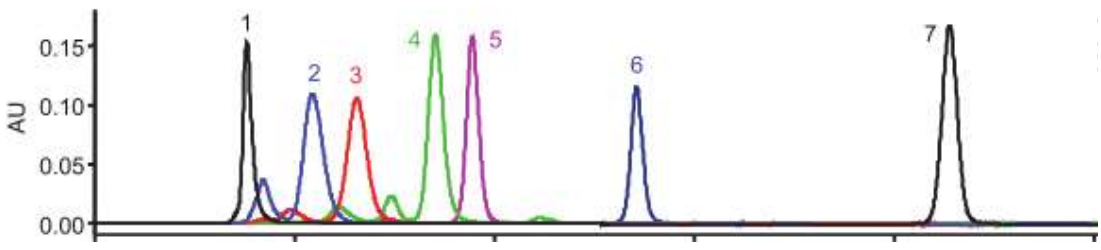

Figure 2-8 SEC separation of model proteins using UPLC. 1) Thyroglobulin 2) BSA 3) Ovalbumin 4) Carbonic anhydrase 5) Myoglobin 6) Angiotensin Fragment 1-7 7) Uracil (Waters Corporation, 2017).

to 22 for BSA and Ova for which baseline resolution was barely achieved using a UPLC with a small bead size.

2.5.3 Analytical separation using ion exchange HPLC

Ion exchange HPLC separates solutes based on charge. In AEX-HPLC the solute is negatively charged, this negative charge is created by using a suitably high pH in the mobile phase. The stationary phase is positively charged and interacts with the solute via electrostatic interactions. The solute can then be eluted either by weakening the interaction between solute and stationary phase by decreasing the pH in the case of AEX-HPLC, or by increasing the concentration of competing ions such as Cl^- which will displace the solute from the stationary phase. Cation exchange chromatography works via the same mechanism except positively charged solutes are exchanged on a negatively charged stationary phase (Karlsson & Hirsh, 2011).

Investigation of analytical separation of BSA, Ova and Con on a strong cation and strong anion exchangers has previously been done (Kopaciewicz, Rounds, Fausnaugh, & Regnier, 1983). Displayed in Figure 2-9 are the retention times of various model proteins including BSA, Ova and Con on strong anion and cation exchangers using the same linear NaCl gradients across all datasets. First strong cation exchangers will be considered: Con albumin retention times increase at pH 7 and below, BSA retention increases at pH 5.5 and Ova increases at pHs 4.5. This means that the highest pH at which all 3 proteins are showing significant retention is pH 4.5. The line crossing the BSA cation plot at pH 5.5 represents the pH at which BSA starts showing poor recovery, at pHs tested below 5.5 BSA showed poor recovery, although the authors present no explanation as to why recovery worsens. Taken together this means that for strong-cation exchangers there is no pH at which BSA, Ova and Con show both retention and good recovery.

Now consider the data for strong anion-exchangers: It is clear that at pHs above 6 BSA, Ova and Con all show retention meaning AEX-HPLC may show some

promise for analytical separation. AEX-HPLC chromatograms for the model proteins is displayed in Figure 2-10. This data was generated using a monolith analytical column which attempts to maximise resolution by minimising peak width and band spreading through the use of short diffusion paths (Ersson, Rydén, & Janson, 2011). The chromatogram shows significant overlap between BSA and Ova species as well as some overlap between Ova and BSA. From examination of Figure 2-9 it is clear that the highest pHs tested show the greatest difference in retention time between BSA and Ova as increases in retention tail off for Ova but continue to increase in the case of BSA. At pH 9 significant overlap between BSA and Ova is still observed meaning that without some sort of deconvolution method analytical quantification using this method is not tractable for all the binary and ternary isotherms.

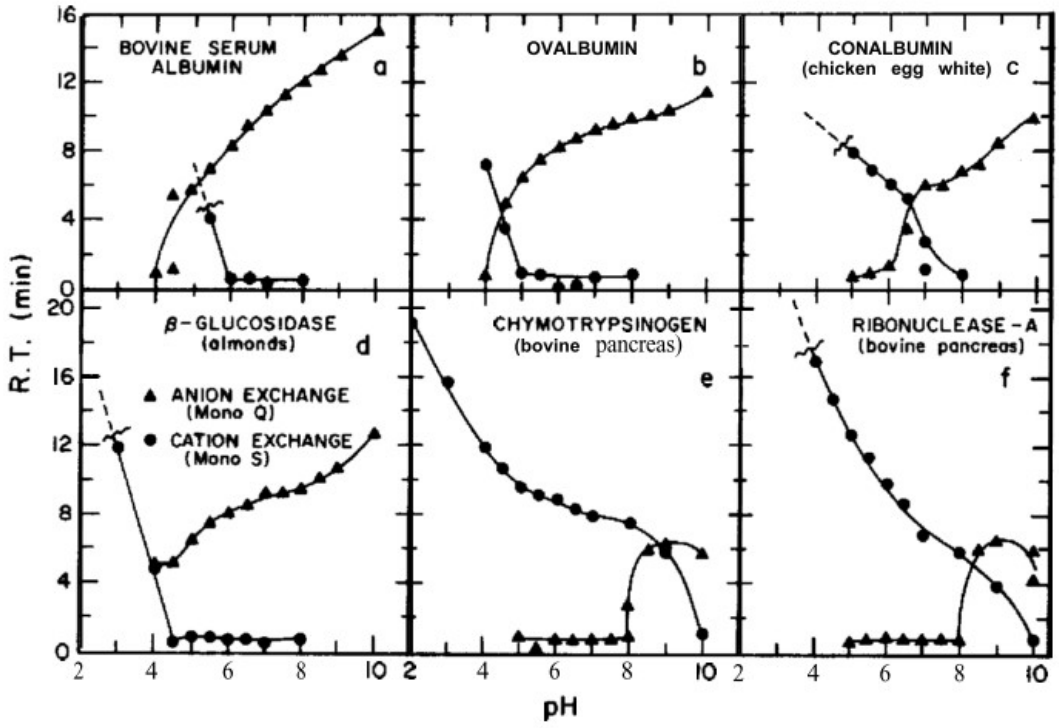

Figure 2-9 Investigation of the effect of pH on retention time using strong anionic and strong cationic exchangers using the same linear NaCl gradient. Figure has been reproduced from (Kopaciewicz et al., 1983).

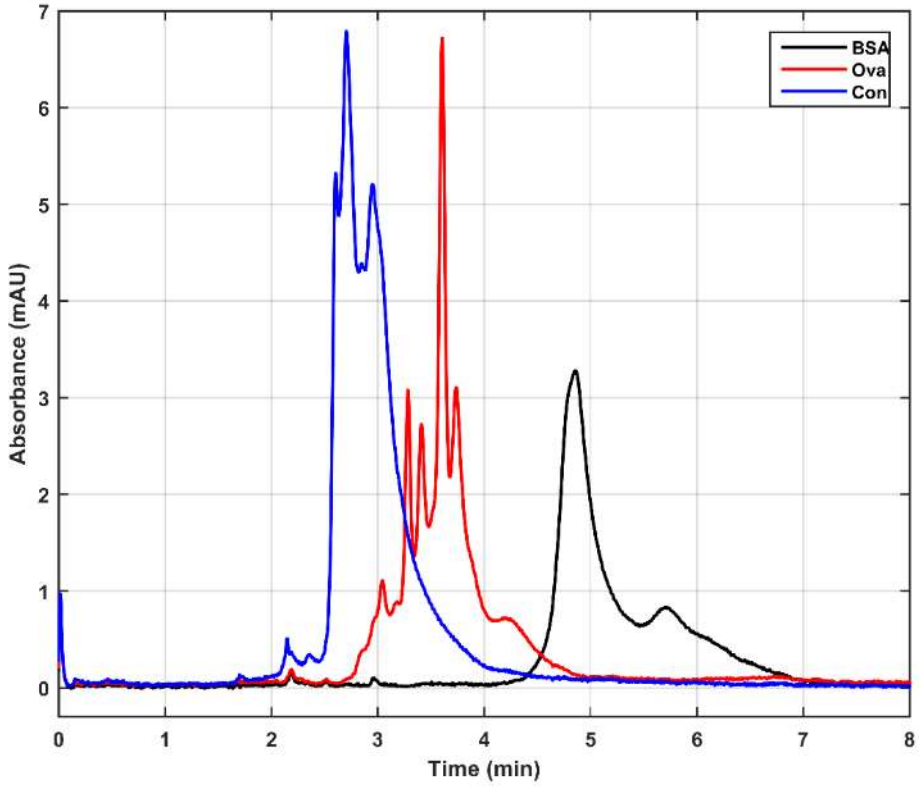

Figure 2-10 Analytical anion exchange HPLC chromatogram. BIA CIMacTM monolith 0.1 mL column 0-1 M NaCl in Tris pH 9 gradient 75 mM/min at 1.5 mL/min. Multiple Con peaks are due to Con carrying different levels of Iron, Ova peaks are caused by differentially phosphorylated Ova species and multiple BSA peaks are caused by the presence of monomer, dimer and aggregate.

2.5.4 Summary of HPLC methods

As mentioned previously baseline separation of all three proteins is needed to generate a multicomponent isotherm free of systematic error. Alternatively, a good separation in conjunction with an effective deconvolution method could also be effective. The analytical methods explored here do show some promise of achieving baseline separation for at least 2 components such as HIC separation of BSA from Ova and Con or UPLC SEC separation of Ova from BSA and Con. Potentially 2 analytical methods run in tandem could be effective. Quantification of BSA in a HIC run in which the co-eluting BSA-Ova peak could be fractionated and run on SEC-HPLC. There are however other factors to consider for an ideal analytical method.

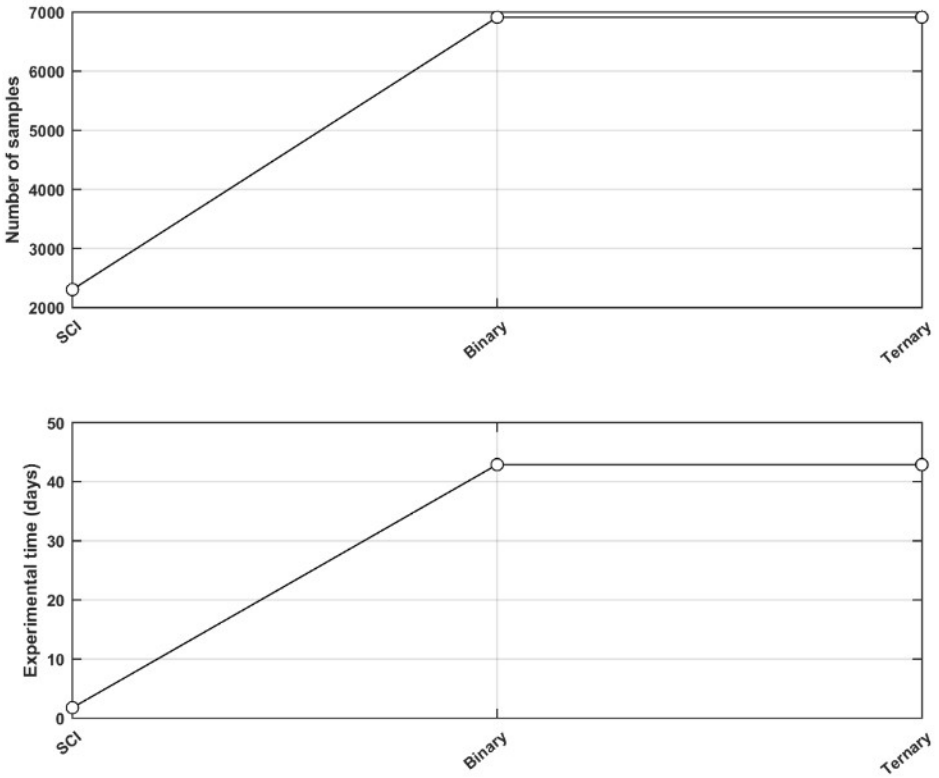

Figure 2-11 Number of analytical samples generated and experimental time taken for such studies. The assumptions for the number of samples generated are that 2 pHs are being studied at 3 NaCl levels for 3 proteins with each load having 2 replicates, each load would go through elution cycles meaning equilibrium samples and 3 wash/elution samples for each load would be analysed. One PreDictor could run 3 single component isotherms or one binary and a ternary would require 3 PreDictor plates. Assumptions for time taken are that one PreDictor plate taken 6 hours to run to elution, SCI analysis takes 9.6 seconds to analyse 96 samples at 280 nm and 8 minutes to run HPLC on one sample.

Generation of multicomponent isotherms requires the creation of many samples. In a binary many combinations of loads are required to explore the displacement of one protein for another as well as having enough points to explore the shape and curvature of the isotherm, additionally isotherms need to be generated at a range of buffer conditions exploring the suppression of the binary by the buffer modulator which weakens interaction between solute and sorbent to cause elution. In the case of ion exchange different pH levels could also be explored which requires the exploration of different loads and different NaCl levels again. Additionally there are 3 binary isotherms to potentially explore as well as the ternary isotherm which requires even more load combinations not to mention its exploration

and different NaCl and pH levels. Moreover all protein loads at all pHs and NaCl levels should be run at least in duplicate to have a reliable data set. Figure 2-11 explores the number of samples required and the experimental time associated with carrying out an isotherm study for one 3 component system using HPLC analysis, the number of samples required for 3 binary or 1 ternary isotherm is the same because there are 3 binary combinations with a smaller number of samples in each vs 1 ternary with an increased number of samples. Figure 2-11 shows it would take approximately 90 days to run one three component study at 2 pHs. Figure 2-11 only includes experimental time and does not include planning time nor time taken to integrate, or time involved in the deconvolution of HPLC chromatograms. It also assumes an 8 minute HPLC run time as suggested by Figure 2-10 and not a complex 2 step analytical separation described earlier. In summary not only is HPLC separation of the model mixture challenging it is also prohibitively time consuming. Ideally an alternative method for quantification would quantify the proteins using a rapid, high-throughput (HT) methodology.

Such a method exists via the collection UV spectra and multivariate data analysis. UV spectra can be analysed in a high throughput format using 96 well UV plates and an appropriate plate reader. Spectral measurement of a set of mixtures of known concentration can then be used to calibrate a multivariate statistical model which links the variation in spectra to the variation in concentration of each component. The spectra of samples of unknown content which lie in the space of the model can then be supplied to the model which will then return estimates on the concentration of each component (Hansen, Jamali, & Hubbuch, 2013; Hansen, Skibsted, Staby, & Hubbuch, 2011; Kamga, Woo Lee, Liu, & Yoon, 2013; Rathore, Li, Bartkowski, Sharma, & Lu, 2009). The model mixture of BSA, Ova and Con has not been published before in a multicomponent study: The increased molecular weight, lack of a chromophore and chemical similarities of this mixture should mean it is a challenging mixture to study with relevance to real world problems.

Quantification of protein mixtures using UV spectra is an interesting alternative to HPLC. Whilst there is significant overlap in the spectra of BSA, Ova and Con which will require a deconvolution method in order to back out individual concentrations from a mixture the analysis time required to generate a UV spectra for 96 samples is approximately 30 minutes making it a rapid analytical method well suited to multicomponent isotherm studies. Different proteins will have different UV spectra dependant on the presence of aromatic residues such as tryptophan, tyrosine, and phenylalanine which have a local maxima around 280 nm. Additionally there is another local maxima around 230 nm due to the aliphatic chain as displayed in Figure 2-12 (Hansen et al., 2011). The multivariate model which can be used to link variation in UV spectra to protein concentration variation is partial least squares regression (PLS) which will be explored next.

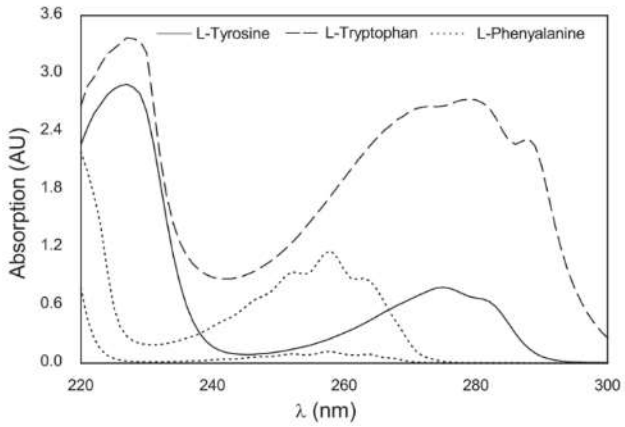

Figure 2-12 UV absorbance spectra of aromatic amino acids (Hansen et al., 2011).

2.5.5 Partial least squares regression overview

The aim of the PLS model is to link variation in the spectra to variation in the concentration of individual components in a mixture. This is done by creating a set of control samples which have known concentrations of all components and measuring their spectra, this information is then used to calibrate and validate the PLS model (Sjöström, Wold, Lindberg, Persson, & Martens, 1983).

Before model building both the X data (spectra) and Y data (concentrations of individual components) are scaled. In this work all data sets were scaled to unit variance scales but not mean centred, work not shown and other published work suggests that there is little consequence of using mean centring or mean with unit variance scaling for this specific application (Kamga et al., 2013).

PLS can be explained in geometric terms which allows one to visualise how models are constructed. Imagine a situation where an observed data set has N observations with 3 X variables associated with it and 1 y variable (in the case presented here that would correspond to 3 wavelengths and 1 protein concentration, additionally y is left lower case as there is only 1 y observation). Each observation can be observed as 2 data points, one in the X space and another in the y space creating 2 swarms of data points, 1 in each space as shown in Figure 2-13. The goal of the analysis is to describe the relationship between the position of the points in X space to their position in y space. Note that in the examples given below data has been mean centred and had unit variance scaling applied.

Once the data has been scaled and plotted the first component or latent variable (LV) can be calculated. The first component is a line which approximates the X data swarm and correlates well with the y data shown in Figure 2-14. The co-ordinate of the observations along this line is found by projecting the observation onto the line. This co-ordinate is known as the score t_1. The estimate of y, ($\hat{y}$) is calculated by multiplying the t_1 score by the weight of the y vector C_1, see Figure 2-14. The differences between y and $\hat{y}$ are the residuals and can be described as the unexplained variation in y after the first latent variable, this represented by the scatter of the points around the diagonal in Figure 2-14.

The descriptive power of the PLS model can be improved by the addition of another LV. Additional LVs are added in the X space and represented as an additional line in that space which passed through the origin and is orthogonal to the first LV. The second LV is chosen to improve the description of the X data as much as possible whilst providing good correlation with the y residuals f_1 after the first LV. The score

of the observation t_2 is calculated by its projection onto the second LV line. t_2 multiplied by the second weight c_2 of the y data gives an estimate of f1, the residual from the first latent variable. The right hand panel of Figure 2-15 is similar to the right hand panel of Figure 2-14. So far only the relationship of t_1 and t_2 and the y data have been considered one at a time. The combined power of LVs 1 and 2 can also be considered in predicting y. An estimate of y after calculating 2 LVs ($\widehat{y_{(2)}}$) by computing $C_1 t_1 + C_2 t_2$.

The number of LVs to include in a model is a pertinent question, enough LVs should be chosen to capture enough of the signal in the X data set in order to predict the y value with acceptable error. However, any calibration dataset will additionally contain noise as well as signal in both the X and y blocks, increasing the number of LVs will lead to the model increasingly describing the noise rather than the signal leading to the model performing more poorly on external data sets with different noise profiles. The number of LVs in a PLS model is chosen using a process called cross validation. In cross validation a portion of the calibration data set is left out of model building, that subsequent model is then used to predict the concentration of the omitted data. From this a measure of the error in the model is acquired at that number of LVs which is the root mean squared error in cross validation (RMSEcv). Ideally, as more LVs are added the RMSEcv is expected to decrease as more of the signal is captured, beyond a certain point the additional LVs are expected to pick up the noise in the data resulting in an increase in the RMSEcv, the optimal number of LVs in this scenario lies in the minimum RMSEcv. An alternative scenario is that there is very little noise in the data set and the addition of further LVs has very little effect on the RMSEcv, in this scenario one would normally pick the smallest number of LVs which gives acceptable RMSEcv (Eriksson, Byrne, Johansson, Tyrgg, & Vikström, 2013; Wold, 1978).

Once model building is complete an additional control data set of known concentration should be used to assess the performance of the model. Ideally this data set should perform with similar levels of error to the calibration dataset, this process is called validation (Eriksson et al., 2013).

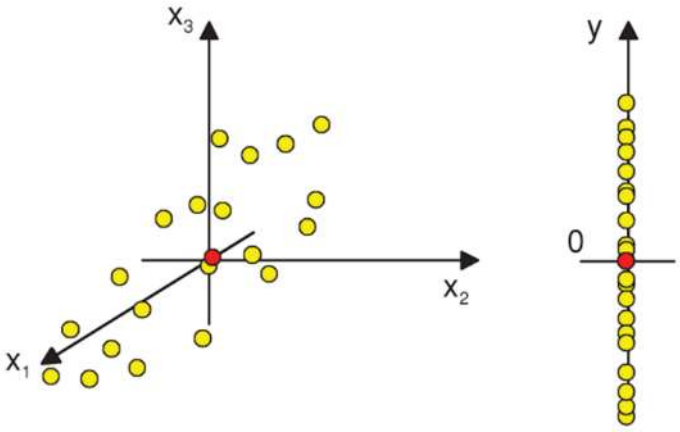

Figure 2-13 Visualisation of X and y data points before PLS model building figure taken from (Eriksson et al., 2013).

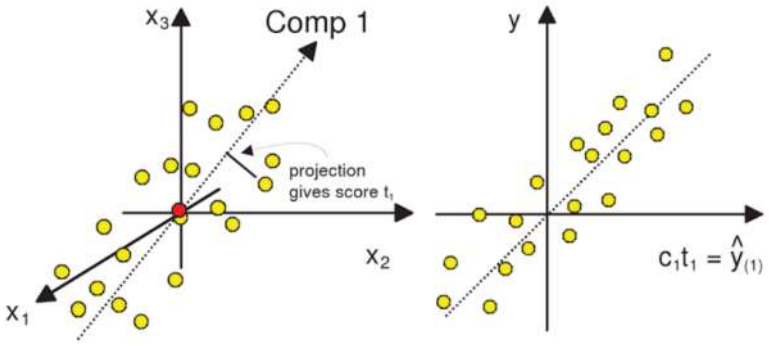

Figure 2-14 The first component drawn in the X and y spaces taken from (Eriksson et al., 2013).

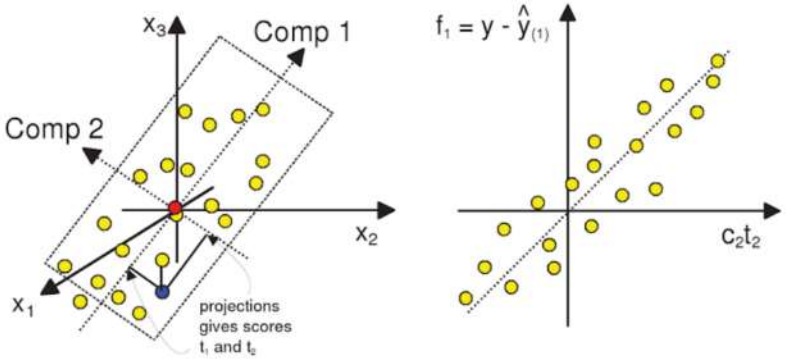

Figure 2-15 Representation of drawing the second latent variable.

2.6 Summary

Biopharmaceuticals are of great and increasing importance both in the treatment of disease as well as in financial terms. Modern manufacturing canonical processes produce API in suspended cell culture, which could either use microbe such as *E. coli* or using immortalised mammalian cell lines. Following recovery via centrifugation and/or filtration, downstream processing is then applied to the resultant solution. Downstream processing removes unwanted impurities and involves application of preparative chromatography column steps. The behaviour of a solute during chromatography is dependent on the isotherm which describes the interaction of the solute and sorbent at equilibrium. Typically these steps often involve loading columns close to their maximum capacity, under these quasi-overloading conditions the multiple solutes in the feed are in competition with one another for binding on the sorbent. As such their elution behaviour is dependent on multicomponent isotherms. Multicomponent protein adsorption isotherms have received limited study historically due to the laborious nature of their elucidation involving many protein load combinations of solute as well as the study of multiple buffer conditions. As such many have previously attempted to use parameters obtained from comparatively facile single component isotherms to the multicomponent cases with varying degrees of success.

Advances in high-throughput chromatography formats have yielded slurry plates which are typically 96 well plates. In every well of a slurry plate there exists a small volume of sorbent which can be loaded with liquid phase and solute representing a miniaturised, parallelised version of the traditional batch adsorption experiment. In conjunction with ALHS this format can facilitate the study of both single and multi-component isotherms. Moreover, the availability of software packages eases the use of multivariate data analysis facilitates the possibility of using UV data in order to assay the concentrations of individual proteins within a mixture. This approach further expedites the study of multi component isotherms as the analytical bottleneck associated with high throughput chromatography formats is alleviated (Konstantinidis, Kong, & Titchener-Hooker, 2013). Furthermore, a

literature search reveals the analytical separation of the model mixture explored here for isotherm studies is challenging meaning the need for an alternative analytical method is pressing.

During completion of the work presented in this thesis another group successfully demonstrated the application of slurry plates using UV spectra and multivariate data analysis in the elucidation of multicomponent isotherms (Baumann et al., 2016). Whilst this work represents an important proof of concept in the area there are some factors which limit the general applicability of the work. Firstly the only binary isotherm study was Cyt-C and Lys. Cyt-C contains a heme group providing an additional chromophore at around 400 nm which not only offers the possibility of facile spectrophotometric quantification of the binary mixture but also changes the UV spectra of the protein easing multivariate quantification of the mixture (Hansen et al., 2013). As most proteins contain no such heme group or additional chromophore the work does not fully demonstrate the application of the method to more realistic problems. Additionally, both Cyt-C and Lys are small proteins (~13 kDa) meaning their behaviour on isotherms is likely to be less confounding than larger proteins.

The HT technologies, ALHS and availability of multivariate data analysis software packages mean that a new opportunity exists in the study of multicomponent protein adsorption which will be explored in this thesis.

2.7 Aims and objectives of thesis

Adsorption isotherms in chromatography are critical in determining separation during column runs (Schmidt-Traub, 2005). Single component protein adsorption isotherms have been extensively studied in the literature but have limited applicability for preparative separations which are often performed under overloading conditions. The batch adsorption experiment format is best placed for the study of multicomponent protein adsorption isotherms. Advances in high-throughput (HT) chromatography formats have yielded slurry plates which represent a parallelised scaled down version of the batch adsorption experiment. Slurry plates are often used in conjunction with automated liquid handling systems (ALHS) which bring the potential advantages of automation, increased throughput, as well as improved accuracy and repeatability.

As mentioned previously, multi-component adsorption isotherms have not been extensively studied, much of the work that has been done involved either small proteins such as lysozyme, or included proteins which contain a chromophore such as cytochrome-C (Cyt-C) (Baumann et al., 2016; Cano et al., 2005; Demin et al., 1997; Garje et al., 1999; Rege et al., 2003; G L Skidmore & Chase, 1990; Xu & Lenhoff, 2009). The presence of a unique chromophore such as in the case of Cyt-C facilitates the quantification of these mixtures and makes the study of their behaviour, not only less error prone but also limits the applicability of the methods to more general cases, which are more relevant as very few therapeutics contain unique chromophores. Additionally the limited size and complexity of these model proteins potentially means their chromatographic behaviour is less complex in comparison to larger, more industrially relevant proteins. As an extension it is feasible that the competitive behaviour of these smaller, more simple proteins is easier to capture using the commonly applied isotherm formalisms and that the problem of studying multicomponent adsorption isotherms has been over simplified in such problems, when compared to larger proteins which are closer to industrially relevant macromolecules.

Multicomponent adsorption isotherms are known to be difficult to study and their elucidation is often avoided or minimised in favour of using parameters taken from single component studies and applying them to limited multicomponent data sets, with some successes (Faraji et al., 2015; Fargues et al., 1998; Rege et al., 2003; Tao et al., 2011; Zhou et al., 2007) and some failures (Aboudzadeh et al., 2007; Finette et al., 1997; Lewus & Carta, 1999; Liang et al., 2012; G L Skidmore & Chase, 1990; Xu & Lenhoff, 2009; S. Zhang & Sun, 2003). It has been suggested that multicomponent isotherm elucidation is challenging, not only due to the sheer quantity of data required, but also because they are error prone (Xu & Lenhoff, 2009).

The aim of this thesis is to study multicomponent isotherms using modern, rapid HT chromatography and analytical techniques. The proteins being studied are BSA (67 kDa, pI = 4.8), Ova (45 kDa, pI = 4.5), Con (77 kDa, pI = 6.6) on the strong anion exchanger Capto^{TM} Q. These proteins are larger than many of the proteins typically studied which are around 20 kDa. The errors associated with generation of single component isotherms has been studied (a. Osberghaus et al., 2012) but the propensity for error associated with multicomponent study, which is one of the reasons they have not been extensively studied, remain poorly understood. As such the systematic and random errors associated with these systems and ways to minimise these errors will be a central topic.

Initially critical experimental variables in the generation of low error single component isotherms will be identified. As well as methodologies to generate data with minimised error. Following that single component isotherms will be elucidated for all 3 protein across different pH and NaCl levels.

Once a HT platform for the study of isotherms has been established, an analytical method for the quantification of protein mixtures will be developed and presented. Typically an HPLC method would be chosen and optimised for such a task. However, extensive research into the literature suggests that no single HPLC method would give acceptable resolution without the use of deconvolution methods. As such an alternative analytical method is deployed: The method involves the rapid, HT

collection of UV spectra which is used to quantify protein mixtures in conjunction with a previously calibrated multivariate model.

Using the optimised HT platform for the generation of isotherms, with the rapid HT analytical method for the quantitation of proteins, binary and ternary isotherms are produced across pH and NaCl levels. Additionally, the propagation of systematic and random errors through to adsorbed concentration is performed. Ternary adsorption isotherms for proteins are very rare and limited studies have only been reported on a few occasions (Close et al., 2014; Melter et al., 2007; Tao et al., 2011).

The resulting multicomponent adsorption isotherms will then be fit to different isotherm formalisms to see which, if any, are able to fit the complexity of these isotherms across different levels. The optimisation of a BSA-Ova separation on columns will then be investigated.

The overall aim of the thesis is to identify and optimise methods associated with the study of multicomponent protein adsorption isotherms which have received limited study historically.

Chapter 3 Materials & methods

3.1 Generic materials used across methods

BSA (pI 4.8, cat no. A2153, lot no. SLBC3307V), ovalbumin (Ova, pI 4.5, cat no. A5503, lot no. SLBD2312V), conalbumin (Con, pI 6.6, cat no. C7786, lot no. SLBF7194V) and Bovine haemoglobin (Hb, pI 6.8 cat no. H2500, lot no. SLBF3496V) were purchased from Sigma Aldrich (Gillingham, UK) in the form of a lyophilised powder, proteins were reserved from one batch to avoid issues around batch to batch variability.

The automated liquid handling system (ALHS) used in this work is the Tecan Evolution 100 (Tecan, Mannendorf, Switzerland) fitted with fixed tips, a robotic manipulation arm to move labware from location to location, 1 mL diluters, a TeShake™ and a Tecan Infinite® 200 plate reader unintegrated plate reader as displayed in Figure 2-1. The ALHS was controlled using Evoware 2.3® (Tecan). Tris stock was made by dissolving Tris base and Tris-HCl to a final concentration of 1 M in a predetermined ratio to achieve a pH of either 8 or 9 in ultra-high quality water (UHQ) (Millipore, Bedford, USA). This stock was mixed with other stocks when required and was used to make all Tris based solutions after dilution with UHQ water. NaCl and NaOH solutions were prepared in a similar way. All salts were at least analytical grade in quality.

3.2 Sanitisation and carry over testing during liquid handling

The reason for using an ALHS fitted with fixed tips rather than one set up to used disposable tips is discussed in Chapter 2. The use of a fixed tip requires tips to be washed and possibility sanitised of any solution after liquid handling to ensure no carry over. Additionally, the removal of any sanitisation solution used also needs to be validated to ensure there is no contamination of solutions handled during experimentation.

Lyophilised proteins were dissolved in 5 mM Tris pH 9 at a concentration of 20 mg/mL. The sanitisation solution used to remove any residual protein was 0.5 M

NaOH which was achieved by diluting the 1 M stock. The sanitisation procedures tested always involved washing the inside and outside of the tip with a variable volume of ALHS liquid (UHQ water) before aspirating and dispensing 975 µL of sanitisation fluid (0.5 M NaOH) followed by washing the NaOH solution off the inside and outside of the tips using a variable volume of system liquid, once an effective wash scheme was decided upon the wash volumes were fixed. Tip washing could be done slowly using the syringes, or quickly using the fast wash module which is a fast liquid pump. Sanitisation testing was done by requesting the ALHS to aspirate and dispense 900 µL of protein stock twice with a sanitisation procedure after each protein handling. The ALHS then transferred 150 µL of 5 mM Tris pH 9 diluent as a test solution to be checked for protein content. Negative controls were also tested by transferring 150 µL of diluent before the test began, an additional negative control was tested by transferring 150 µL of diluent using a hand operated disposable multi-channel pipette. Positive controls were also prepared by serially diluting protein stocks to 10 and 1 µg/mL, 150 µL of these positive controls was also transferred by hand to a 96 well 300 µL collection (Starstedt, Nümbrecht, Germany cat no. 82.1581. A whole column of the 96 well plate was used for each round of test and control solution evaluation meaning all 8 tips of the ALHS were tested and the number of replicates per protein handling event or control test was 8. 150 µL of Bradford reagent (Thermo Scientific™ cat no. 41116012) was added to the test and control wells using the hand pipette before mixing via pipetting and incubation at room temperature for 30 minutes before being read on the plate reader At 595 nm. The analytical method in question is the micro-Bradford assay which has an enhanced limit of detection in comparison to the standard Bradford assay.

In addition to testing for protein carryover during liquid handling events the carryover of sanitisation liquid (0.5 mM NaOH) was also tested. This was done in a separate test by requesting the ALHS to perform the sanitisation procedure 20 times, after each event the ALHS was requested to aspirate and dispense the same aliquot of UHQ water. This was done so that any small amount of carry over with each cycle of sanitisation would accumulate and facilitate detection. The water aliquot was pH checked using a pH probe (Mettler Toledo, Giessen, Germany) before and after

testing alongside a pH 7 control which provides information on the variability of the pH measurement.

3.3 Gravimetric liquid handling assessment and calibration of experimental solutions

Liquid handling accuracy and precision (described in equations 4-1 and 4-2) of the ALHS was assessed by requesting the ALHS to dispense a predetermined volume of test solution into a vessel which was tared on a balance (AB204-S, Mettler Toledo, Giessen, Germany) immediately before dispensing, vessel and liquid were then weighed and recorded. In order to correct for density a 5 mL pycnometer (Lenz Laborglas, Wertheim, Germany) was used. The pycnometer was filled with test solution and weighed before being cleaned and filled with UHQ water and weighed again. The weight of the test solution was then divided by the weight of the UHQ water. Assuming the UHQ water has a density of 1 g/mL and the pycnometer was filled to the same level across both solutions this calculation will return the density of the test solution without assuming the volume of the pycnometer.

It was first assessed whether the ALHS had similar liquid handling performance across all tips. This was done by gravimetrically assessing 20, 50, 170 and 700 µL of dispensed UHQ water using all 8 tips with 3 replicates per tip. The results of an ANOVA showed p-values were greater than 0.05 for all volumes tested except 20 µL, suggesting the tips behaved the same in this test at 50 µL or above. Once the comparability of the tips was established the liquid handling performance was assessed by requesting the Tecan to dispense 20, 50, 170 and 700 µL of test solution for gravimetric assessment. This was replicated 5 times with the tip being tested rotating with every replicate so that all tips were used for the assessment. If any of the test solutions had an accuracy greater than 1% (equation 4-1) at a volume of 50 µL or more it was taken forward for liquid handling calibration.

Liquid handling calibration was performed in the same format as the assessment except additional volumes were tested. The ALHS liquid handling was divided into volumes ranges (VRs) which are as follows, all expressed in µL: VR1; 19.9-

40.01, VR2; 40.01-95, VR3; 95-350.01, VR4; 350.01-975. Each of these VRs can be calibrated separately by changing the slope and intercept of the VR calibration curve. The VR calibration curve displays delivered volume on the x-axis and desired volume on the y-axis. In order to calibrate the ALHS each VR should contain at least 3 tested volumes, as such the tested volumes in each VR were (µL): VR1; 20, 30, 40. VR2; 50, 75, 95. VR3; 110, 170, 250, 350. VR4; 550, 700, 950. Once the liquid handling accuracy had been found adjustments to each VR calibration slope and intercept were calculated in order to correct for any deviation in accuracy. The liquid handling performance was then reassessed for the solution under calibration. Only one round of recalibration was required in order to achieve liquid handling accuracy within 1% at 50 µL and above, if this had not been achieved the calibration process would have been repeated (Treier et al., 2012).

$$\text{Accuracy} = \frac{\text{Average observed} - \text{Nominal}}{\text{Nominal}} . 100 \qquad \textit{3-1}$$

$$\text{Precision} = \frac{\text{Standard deviation}}{\text{Average observed}} . 100 \qquad \textit{3-2}$$

3.4 High-throughput isotherm data generation

Isotherms are described by data points which have both a liquid phase equilibrium concentration and a stationary phase adsorbed concentration associated with them. In this work two versions of each data point are determined in a batch adsorption experiment. Data type one, which is what has been determined historically, relies on the equilibrium concentration being assayed directly and the adsorbed concentration being calculated via a mass balance using the known starting concentration. Data type two, which is being explored in this work, involves eluting what was adsorbed to the stationary phase and assaying the eluted fractions to independently ascertain the adsorbed concentration, the equilibrium liquid concentration can also be determined independently of data type one by calculating it via mass balance using the starting concentration.

Isotherm data was generated using a specific type of slurry plate called a PreDictor™ plate. A slurry plate is deep 96 well plate in which every well location contains a small volume of sorbent, in the case of the PreDictor™ plate it is also a filter plate allowing easy separation of liquid and solid phases by either centrifugation or vacuum filtration. The PreDictor™ plates (GE Healthcare, Uppsala, Sweden) were filled with Capto™ Q (GE Healthcare, Uppsala, Sweden), resin at volumes of either 2 µL, 20 µL or 50 µL. The described slurry plates and robotic station were deployed for the development of isotherms involving the aforementioned proteins. These were measured at various liquid conditions employing 50 mM Tris buffers at pHs of 8 or 9 and with various NaCl concentrations. A schematic methodology is described in Figure 3-1. The slurry plates were equilibrated and washed using the same buffer composition, whereas for their elution and stripping different buffers were used with higher NaCl concentrations (i.e., of 0.3 M and 1 M respectively). The equilibration, wash, elution and strip buffers were either prepared in 2 mL 96 well deep square well plates (Porvair Sciences, Wrexham, UK) by mixing stocks of Tris HCl, Tris Base and NaCl at the appropriate ratios, or were prepared manually and transferred to troughs.

The slurry plates were loaded with protein feed solutions prepared at the same liquid conditions with the equilibration (and wash) buffers per well. To generate the isotherms, different load challenges were tested in different wells and they were achieved by varying the protein concentration. These feed solutions were prepared on the robotic station by employing the custom liquid classes described earlier. For this purpose, protein stocks were prepared for each protein, at concentrations of 2 mg/mL and 20 mg/mL, by dissolving their appropriate amounts in a 5 mM Tris pH 8 (or 9) diluent buffer. The protein stock solutions, diluent, 1 or 2 M NaCl, and 1 M Tris at pH 8 or 9 were then placed into separate troughs on the deck of the robot. The robot was then instructed to mix the aforementioned solutions in 2 mL 96 well deep square well plates (Porvair Sciences) at the appropriate ratios so as to generate 1 mL of 50 mM Tris feed solutions with the desired pH, protein and NaCl concentration. Protein loads were run in duplicate and after incubation and elution between 1 and 3 different dilution levels of both equilibrium and elution 1 samples were analysed

for multicomponent isotherms, meaning each protein load had been 4 and 12 data points associated with it. These replicates were put forward for outlier rejection screening discussed below.

The filter plates were handled following the manufacturer's instructions with a few changes. In order to properly equilibrate the chromatography resin with the liquid phase the sorbent particles must be suspended in the liquid, this is achieved by vigorous shaking on a plate mixer. All mixing steps took place on a TeShake™ orbital shaker (Tecan), which was integrated into the robotic station. The agitation speed was set to 1300 rpm to ensure that the resin was fully suspended (Bensch et al., 2005) and the plates were sealed with aluminium plate sealers (Greiner Bio-One©, Stonehouse, UK) to prevent cross contamination during agitation and evaporation over prolonged periods of time. The plates were incubated with the protein feed solutions for a duration of 1.5 hrs under shaking. Up to three wash and elution cycles were carried out and here the addition of buffers was followed by 10 minutes of shaking. Finally, the plates were evacuated by centrifugation at 500 g for 1 minute with an Avanti ® J-E centrifuge (Beckman Coulter©, High Wycome, UK). The flowthrough, wash and elution fractions were collected in 2 mL deep square well plates (Porvair Sciences) before analysis, or they were stored at 4°C before analysis.

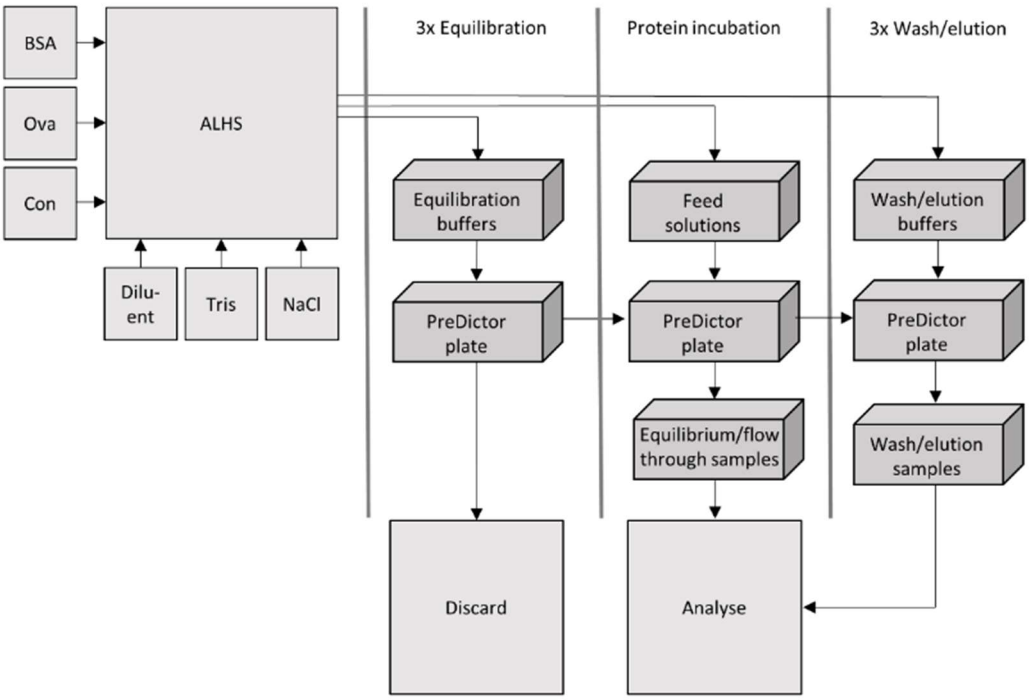

Figure 3-1 Schematic methodology displaying how equilibration, feed, wash and elution buffers are prepared in 96 well plates by mixing different amounts of stock solutions using the automated liquid handling system (ALHS). Bovine serum albumin (BSA), ovalbumin (Ova) and conalbumin (Con) are only used in preparation of the feed solutions. After preparation these solutions are loaded onto the PreDictor™ plate for incubation on the plate shaker before the resulting solutions are either discarded or taken forward for analysis. For multicomponent isotherms multiple dilutions of equilibrium and elution 1 samples have been taken for analysis.

3.4.1 Isotherm value calculations

The notation used in the below equations is generally in line with both historic and more recent publications (Faraji et al., 2015; Graham L. Skidmore et al., 1990). Some of the notation referring to mass is new in this area of study but after a non-exhaustive literature search this is the first work to discuss quantification of isotherms via elution and so new notation is required.

Every well of the slurry plate contains a small volume of chromatography resin (V_{resin} in µL) and a filter. Both the resin and filter are associated with a small volume of liquid (*Vr* in µL, equation 3-3) the supplier, GE Healthcare state in their literature

that the resin retains a ratio of 0.6 relative to its volume and the filter retains 6 µL. This liquid cannot be separated using the centrifugation method described above, therefore every time the liquid in the well is changed there is small carry over of liquid from one step to the next. Starting concentration (C_{start} in mg/mL) must be adjusted to C_0 in mg/mL using Equation 3-4 because the buffer used to equilibrate the resin dilutes the feed, adsorbed concentration calculated via mass balance (q in mg/mL resin) in Equation 3-5 uses the corrected phase ratio F and corrected starting concentration C_0. In historic publications phase ratio has been described as φ which is the reciprocal of F (Conder & Young, 1979; Graham L. Skidmore et al., 1990), F has been used in more recent publication particularly in reference to work using slurry plates (Chhatre & Titchener-Hooker, 2009).

Adsorbed concentration calculated using the second method described earlier where adsorbed protein is eluted and assayed is described in equation 3-7. Adsorbed concentration calculated via elution (q^* in mg/mL) corrects for the protein mass carried over between cycles (m_{ri} in µg) described in equation 3-8. The mass carried over from the equilibrium fraction (m_{req} in µg) to the first wash step and the mass carried over from wash (m_{rw} in µg) or elution (m_{re} in µg) cycles to the next cycle is also corrected for by subtracting it from the total mass in each wash (m_w in µg) and elution steps (m_e in µg). Equation 3-9 describes the second calculation of equilibrium concentration (C_{eq}* in mg/mL) which is a mass balance calculation based on the adsorbed concentration calculated via elution (q*) and the corrected starting concentration (C_0).

$$V_r = 0.6 \times V_{resin} + 6 \qquad 3\text{-}3$$

$$C_0 = \frac{C_{start} \times V_L}{V_L + V_r} \qquad 3\text{-}4$$

$$q = (C_0 - C_{eq}) \times F \qquad 3\text{-}5$$

$$F = \frac{(V_L + V_r)}{V_{resin}} \qquad 3\text{-}6$$

$$q^* = \frac{(\sum_{i=1}^{n} m_{w,i} - m_{rw,i} + \sum_{i=1}^{n} m_{e,i} - \sum_{i=1}^{n-1} m_{re,i} - m_{req})}{V_{resin}} \quad 3\text{-}7$$

$$m_{r,i} = V_r \times C_i \quad 3\text{-}8$$

$$C_{eq}{}^* = C_0 - \frac{q^*}{F} \quad 3\text{-}9$$

3.4.2 Protein quantification

For the quantification of protein concentrations, 100 µL of samples were transferred to half area UV plates (Corning Wiesbaden, Germany) and these were then measured on a Tecan Infinite® 200 plate reader (Tecan). For samples containing a single protein, analysis was done by recording the absorbance at 280 nm whereas for multicomponent samples, spectra were recorded in the 243-300 nm range averaging 10 reads per well. In the former case, the measurements were converted into concentrations via a standard curve whereas in the latter case, wherein multiple proteins were present in a sample, a multivariate analysis approach was employed based on Partial Least Squares (PLS) regression. This was implemented using SIMCA (Umetrics, Sweden) as described below.

3.5 PLS model calibration and validation

The quantification of samples containing more than one protein (multicomponent mixtures) took place via PLS regression. For this purpose, a range of multicomponent mixtures, with known concentrations, were prepared and their spectra were recorded. These protein solutions were made up using the methodology described for the preparation of the feed solutions described in 3.4 High-throughput isotherm data generation: protein combination concentration levels were combined in a space filling grid layout and the corresponding mixtures were prepared in triplicate across 2 mL 96 well deep square well plates (Provair). The measured spectra were blank corrected and the triplicated measurements were screened for outliers using a

median absolute deviation method (see equation 5-1) which is described in further detail below (Rousseeuw & Hubert, 2011). The resultant data sets were employed to build initial PLS regression models, using unit variance scaling which was not mean centred, and their predictions were employed to identify further outliers for removal. The samples in these final data sets were then sorted by their T scores and alternating samples were allocated to calibration and validation data sets. Finally, the resulting calibration sets were employed for calibrating final PLS models in a procedure which implemented cross-validation, as described in (Wold, 1978), to identify the optimal number of latent variables (LVs) by minimising the root mean square error and concurrently minimising the incorporation of noise into the model. The determined models were then employed to return protein concentrations in unknown multicomponent mixtures by supplying the spectra of the unknown samples.

3.6 Outlier rejection screening

Outlier rejection screening was carried out during 2 phases of data processing. Firstly during PLS model building where triplicate spectra were screened for outliers before the remaining spectra where averaged. Secondly during multicomponent isotherm data generation where protein replicate loads were screened for outliers. The *C*, *q* and *q** replicates screened were either true experimental replicates or they represented the same sample diluted to a different level or they represented alternative measures of the same value, i.e. *q* and *q**.

The outlier rejection method used throughout this work was median absolute deviation (MAD). The equations involved in calculating the MAD and score are displayed in equations 3-10 and 3-11. MAD relies on the median values and median deviation around the median instead of using average values and standard deviations. Both the mean and standard deviation can be strongly affected by the presence of just one outlier, whereas the median and median absolute deviation are more robust methods which are not affected by outliers until 50% of the data becomes outlying (Leys, Ley, Klein, Bernard, & Licata, 2013; Rousseeuw & Hubert,

2011). This also means MAD is well suited to outlier detection of small data sets such as the triplicate spectra and isotherm loads which have a minimum of 4 replicates.

In the case of spectra outlier detection the limit for the score was set to 3 which is considered very conservative as opposed to 2 which is considered poorly conservative (Leys et al., 2013). A cut off of 2 would reject spectra which are closer to the median than a cut off of 3. The spectra being considered are from 243 nm to 300 nm at intervals of 1 nm meaning every wavelength in the spectra generates a score, spectra where ≥ 10% of the spectra were considered outlying were rejected and the remaining spectra replicates were averaged. In the case of outlier detection of isotherm data points the cut off score was set to 2.5.

$$\mathrm{MAD} = 1.483 \cdot \mathrm{median}_i(|x_i - \mathrm{median}\ (x_j)|) \qquad \textit{3-10}$$

$$\mathrm{score} = \ (x_i - \mathrm{median}\ (x_j))/\mathrm{MAD} \qquad \textit{3-11}$$

3.7 Kinetic batch uptake method

Kinetic batch uptake experiments in chromatography monitor the uptake of solute, in this case protein, over time in a stirred batch uptake experiment. They only differ in format from equilibrium batch adsorption experiments involved in isotherm study in that samples are taken at time intervals rather than only at equilibrium, if material is removed from the reactor during sample the mass balance calculations involved must reflect that. Traditionally they are used to elucidate parameters and model forms for transport of solute during adsorption. It is possible to run kinetic batch adsorption uptake experiments using an HT format such as PreDictor™ plates (T. Bergander, Nilsson-Valimaa, Oberg, & Lacki, 2008; Traylor et al., 2014). However, bench top beaker scale experiments are more common (Kumar et al., 2014; Tao & Carta, 2008; Wright, Muzzio, & Glasser, 1998). Here beaker scale kinetic uptake experiments will be described.

Capto^{TM} Q (GE Healthcare, cat no. 17531603) was suspended by inversion in the 20% storage solution. From this re-suspended solution 1 mL was transferred to a centrifuge tube and forced to settle via centrifugation. The approximate proportion of resin volume to supernatant was noted down, this proportion was used to calculate the volume of suspended slurry required to achieve 10 mL of sorbent, an extra 30% taken as resin loss is anticipated during preparation. The resin is allowed to settle via gravity in a measuring cylinder before the supernatant (20% ethanol) was removed and 10 column volumes (CVs) of UHQ water are added before the suspension was re-suspended via inversion for 5 minutes. The process of removing the supernatant, replacing it with fresh liquid phase, mixing by inversion and allowing gravimetric settling was repeated 3 times using buffer at the pH and NaCl content required for the uptake study. After 3 rounds of equilibration the supernatant was removed and the slurry transferred to the smallest measuring cylinder available that will accommodate the resin, the slurry was left to settle overnight. Smaller measuring cylinders are preferred as they are associated with less error.

After overnight settling the volume of resin was noted, the volume of resin must be ≥10 mL, the slurry was transferred to a larger measuring cylinder which will accommodate a phase ratio of 5 via addition of liquid phase. The addition of liquid phase to the slurry involves rinsing out the small measuring cylinder to ensure there is negligible loss of slurry during transfer. 50 mL of slurry at a phase ratio of 5 is transferred to the reaction vessel which is a 250 mL beaker, the measuring cylinder is rinsed with a 50 mL of liquid phase which is also transferred to the reaction vessel to minimise resin loss. The slurry was stirred using an overhead stirrer (IKA, Oxford, England) with a Rushton impeller, the suspension is stirred at 50 rpm to maintain full suspension. 50 mL of 60 mg/mL BSA dissolved in liquid phase is also added to the reaction vessel and a timer is started. After all the additions the total volume in the reaction vessel is 150 mL including 10 mL of sorbent. 200 µL samples are taken using a P1000 pipette (the pipette orifice must be wide enough to accommodate sorbent particles) and transferred to a 0.22 µm filter centrifuge tube (Costar, NY, USA, cat no. UY-01937-32) before it is centrifuged at 5,000 g for 30 seconds in order to separate

the solid and liquid phases thus halting the uptake. The resultant liquid phase is held for analysis whilst the solid phase is discarded.

Liquid phase samples taken during the uptake were analysed in the same was during single component isotherm studies as described 3.4.2 Protein quantification and adsorbed concentration is calculated using equation 3-5.

3.8 Propagation of random errors

Random errors were propagated through single component and multicomponent isotherms to ascertain the maximum underlying random error in C^*, q and q^*. Sources of random error and where the value was ascertained are as follows: solute concentration in the liquid, observed experimentally; ALHS precision, observed experimentally; resin volume (V_r), based on (T. J. Bergander, Karf, & Brannstrom-Carlsson, 2008); and feed solution concentration (C_0), based on (a. Osberghaus et al., 2012). Random errors were propagated through equations 3-5, 3-7 and 3-9 using equations displayed in Table 3-1.

Calculation type	*Example*	*Maximum random error*
Addition or subtraction	$x = a + b - c$ *(3-12)*	$\sigma_x = \sigma_a + \sigma_b + \sigma_c$ *(3-13)*
Multiplication or division	$x = \frac{a \times b}{c}$ *(3-14)*	$\frac{\sigma_x}{x} = \frac{\sigma_a}{a} + \frac{\sigma_b}{b} + \frac{\sigma_c}{c}$ *(3-15)*

Table 3-1 Propagation of maximum random error. σ represents random error associated with value. These equations are different to those used to propagate standard deviation as these propagate maximum random error. Care must be taken to stack the variation in such a direction as to generate the maximum propagated error so the precision errors do not cancel.

3.9 Preparative packed bed experiments

All packed bed experiments were performed either on the AKTA purifier (GE Healthcare) or AKTA pure (GE Healthcare) fitted with a UV lamp and adsorption detection at 280 nm. A chromatography column prepacked with Capto™ Q with a volume of 4.7 mL and a length of 10 cm (Hiscreen column, GE Healthcare, cat no. 28-9269-78) was used for all experiments. Single component runs were quantified directly from the 280 nm absorbance by correcting the absorbance observed on the AKTA to the absorbance observed in plate measurements via knowledge of the path lengths. These corrected absorbance units were then applied to the relevant standard curve prepared on UV plates. Multicomponent runs were fractionated and quantified using PLS as described in 3.4.2 Protein quantification. Run variables such as protein load, flow rate, gradient start point, gradient end point and gradient slope were adjusted across different runs.

3.10 HPLC analysis

All HPLC analysis was carried out on an Agilent 1200 system fitted with photodiode array set to collect absorbance values at 280 nm (Agilent Technologies, California, USA) using a 0.1 mL monolith column (BIA separations, Ljubljana, Slovenia, cat no. 110.5113-1.3). Buffer A was 20 mM Tris pH 9 and buffer B was 20 mM Tris pH 9 with 2 M NaCl. Injection was performed at 0 mM NaCl and the gradient was run to 1 M NaCl at a gradient of 75 mM/min and a flow rate of 1.5 mL/min. After completion of the gradient a 2 M strip was carried out before re-equilibration at 0 mM NaCl prior to the next injection.

3.11 Fitting experimental data to isotherm descriptions

Freundlich-Langmuir-Jovanovic (FLJ), SMA and Padé isotherm descriptions were parameterised using codes written in Matlab by Dr. Spyridon Konstantinidis but were run by myself. All equilibrium liquid and adsorbed concentrations were converted into mM concentrations for parameterisation before being converted back into mg/mL for display and discussion. Initial isotherm parameters were found using a differential evolution algorithim (DE) (Price, Storn, & Lampinen, 2005), those parameters were taken forward for further optimisation using a simplex based method using the Matlab function fminsearch (Lagarias, Reeds, Wright, & Wright, 1998). Additionally, the fminsearch algorithm is used to find the parameters from a different starting set of parameters. Two sets of parameters are arrived at; one using a DE followed by an fminsearch and second using only an fminsearch. The parameters with the smallest sum of squared error is chosen as the best set of parameters. Multiple searches were carried out in order to maximise the possibility of finding the global optimum in an attempt to avoid local optima.

Chapter 4 Critical experimental variables affecting isotherm data generation and single component isotherms

4.1 Aims and objectives of chapter

The overall aim of this chapter is to establish foundational understanding and confirmation of methodological practices involved in the generation of single component isotherms (SCIs). Once these factors have been established, SCIs for the three model proteins Bovine serum albumin (BSA), ovalbumin (Ova) and conalbumin (Con) at a range of NaCl levels and pHs will be presented and discussed. Bovine hemoglobin (Hb) will also be assessed in some studies presented in this chapter. The specific aims of this chapter are to:

- Characterise and confirm effective cleaning of fixed pipette tips on the automated liquid handling system (ALHS) as well as discus some of the trade-offs associated with the choice of fixed tip vs disposable tip usage.
- Compare results generated on the high throughput (HT) plate microscale format to those generated using larger bench scale formats.
- Compare data generated using different volumes of resin and different phase ratios within pre-dispensed slurry plates and decide what resin volume to use for future experiments.
- Assess and compare propagation of errors in single component isotherms using different experimental parameters.
- Assess liquid handling capability of ALHS and explore the link between resin volume, stock concentration and liquid handling performance including how issues in liquid handling performance can be dealt with.
- The choice of model proteins will be discussed in terms of their retention on columns.
- SCIs will be presented at 2 pH levels and multiple NaCl levels.
    - Adsorbed concentration in isotherms will be quantified by assaying equilibrium concentration and calculating via mass balance as well as assaying eluted fractions.

- Adsorbed concentration assayed via equilibrium and elution will be compared for the SCIs generated.

The establishment of experimental methodologies which generate data maximising accuracy and minimising variability is critical and identifying experimental parameters which generate reliable isotherm data is essential. At a glance liquid handling reliability may seem elementary but confirmation of its behaviour and exploring the boundaries where it breaks down is a pre requisite for generating high quality reliable data. Additionally, understanding factors that affect data quality is also very useful when going forward with more complex multicomponent systems. SCIs will form the jumping of point from which multi component isotherms (MCIs) can be generated, their elucidation is important firstly from a practical perspective in order to estimate their saturation capacity and sensitivity to NaCl before moving forward to MCIs which are more difficult to study. Additionally it is likely that isotherms determined via equilibrium elution will have different properties and data quality will vary between the two methods under different conditions, the determination of which method works best under what binding conditions is best studied under the less analytically challenging conditions if the SCI rather than the confounding conditions of an MCI. Indeed, it is hoped that such information gleamed in this chapter may help improve the quality of the MCIs in subsequent chapters.

4.2 Choice between disposable or fixed tip and their sanitisation

4.2.1 Fixed tips or disposable tips

The ALHS used throughout this work is the Tecan Evo® 100 pictures in Figure 4-1 which can be equipped with either disposable (DiTi) or fixed tips. The choice is associated with a number of trade-offs which are relevant to the applications in this thesis:

1 Cost: DiTis are disposable and it is recommended they only be used once, implying that every cycle of liquid handling has an associated cost. These costs can escalate rapidly as the number of pipetting cycles required for the use of one slurry plate is high. On the other hand, fixed tips are permanently attached to the ALHS meaning there are no consumable costs with each cycle of liquid handling.

2 Time: Aspiration and dispensing of biological samples such as protein solutions is expected to require extended wash or sanitisation cycles when using fixed tips; such cycles are not needed when using DiTis. When in combination with the large number of pipetting cycles required when handling slurry plates, any additional time taken for the handling of all proteinaceous solutions can be significant (Fregeau, Yensen, Elliott, & Fourney, 2007).

3 Deck space: As previously mentioned the ALHS in question is the Tecan Evo 100 meaning the deck width is 100 cm which is the second smallest deck size available. Space on the deck is taken up with a TeShake™ module required for agitating plates in addition to space for the slurry plate as well as associated deep square well plates and analytical UV plates, meaning space on the deck as at a premium.

4 Liquid handling performance: Fixed tips are associated with improved liquid accuracy and precision as well as the ability to handle smaller volumes with greater confidence (Felton, 2004; Fregeau et al., 2007).

It has been suggested that fixed tips do not suit applications involving biological samples as they tend to be “gooey and sticky” except in cases where highly purified mixtures are used (Felton, 2004). As pre-purified feed stocks are being used here the use of fixed tips is not precluded. Taking all of the above factors into account, fixed tips are the best choice for the application in question as it will minimise costs, keep maximal deck space available and maximise liquid handling. Washing and sanitisation procedures when handling protein containing solutions will increase time taken to run a slurry plate and decrease throughput, but this is an acceptable trade off.

4.2.2 Validation of fixed tip sanitisation

As fixed tips are the preferred choice for the studies to be undertaken, the cleaning and sanitisation processes involved when handling protein solutions is to be established and validated. Although solution carry-over must be negligible, it is desirable that the sanitisation procedure is as fast as possible to reduce experimental time and maximise throughput as numerous protein handling steps are required.

Liquid handling on the Tecan involves creating an air gap between the system liquid, (UHQ water) that fills the ALHS and the sample being handled to prevent the sample being diluted during handling. Tip sanitisation involves washing the tip with water: Tip washing can either be done using the conventional syringe which is slow but regenerates the air gap, or using a fast wash module pump which can quickly move larger quantities of water but does not regenerate the air gap. Tip sanitisation after protein handling consists of the following steps:

1 Water wash on inside and outside of tips to remove excess protein and avoid contamination of NaOH sanitisation fluid.
2 Aspiration and dispensing of 0.5 M NaOH solution to remove any traces of protein.
3 Water wash on inside and outside of tips to remove any carry-over from NaOH to the next solution to be handled.

Protein solutions tested for carry-over were BSA, Ova and Hb at 20 mg/mL dissolved in 5 mM Tris pH 9, because previous pilot experiments had shown 20 mg/mL protein solution were sufficient for isotherm studies. Initially a minimal first wash procedure of 2 + 3 mL (inside + outside tips) was tested, but under this regime the NaOH sanitisation solution became foamy during testing suggesting the initial wash was not effective. Presence of foam in the sanitisation trough is to be avoided as it will interfere with ALHS liquid detection and may compromise subsequent rounds of sanitisation. After a multiple iterations were tested the minimal initial wash to avoid foaming was found to be 10 inside + 10 outside mL of fast wash followed by 5 inside + 5 outside mL slow wash.

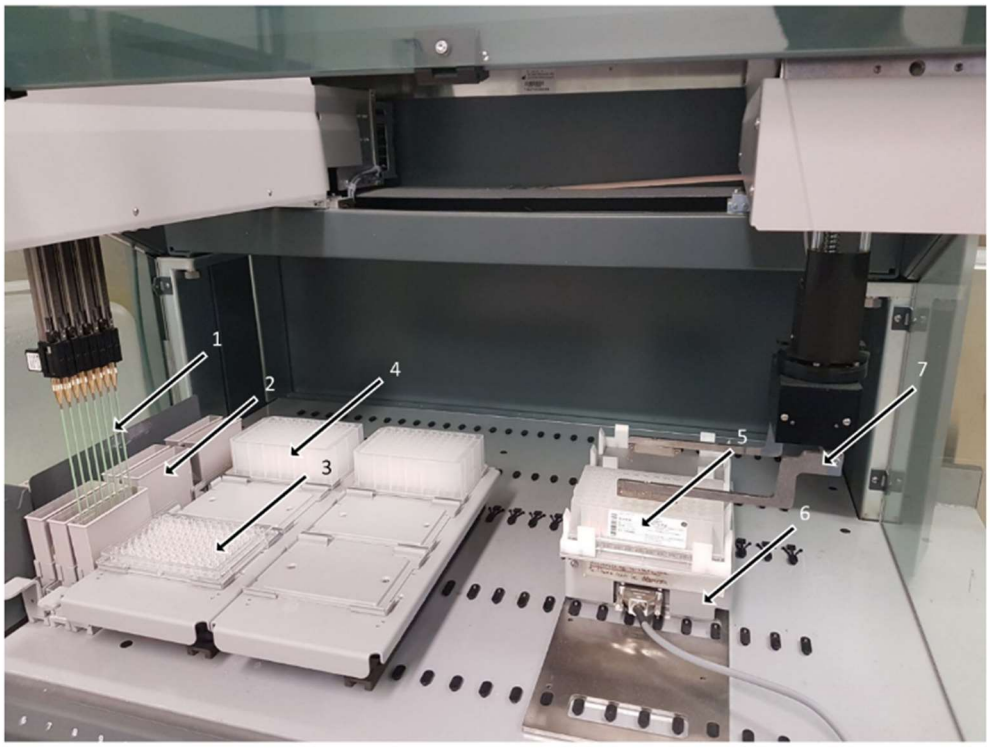

Figure 4-1 Annotated picture of Tecan Evo 100 used in this work. 1) Fixed pipette tips 2) Reagent trough carrying stock solution 3) 96 well UV plate for analysis 4) 96 deep square well plate for preparing equilibration, feed, wash or elution buffers 5) PreDictor plate 6) TeShake for agitating plates 7) Robotic arm for moving plates between locations.

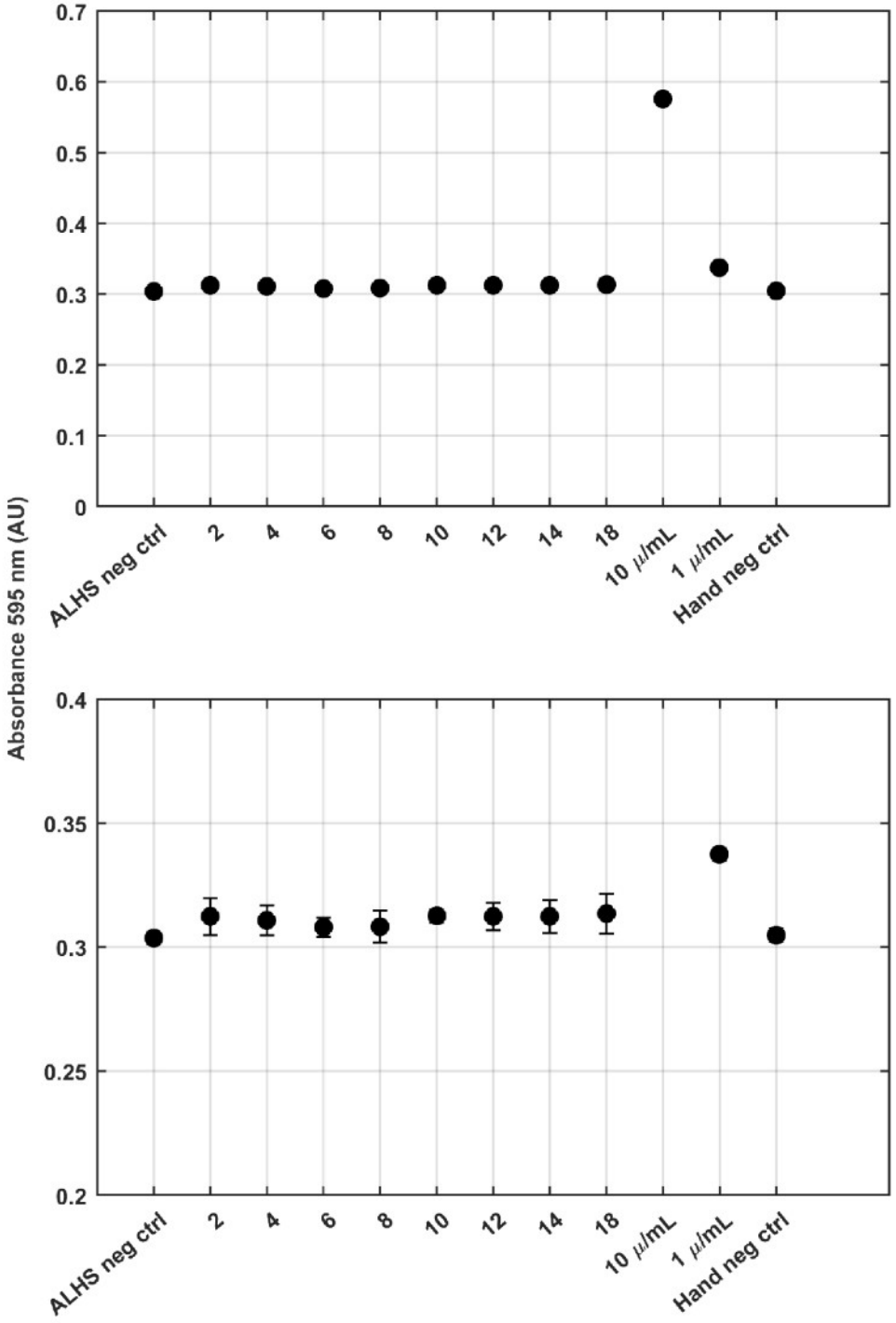

Figure 4-2 Micro-Bradford assay results for sanitisation of Hb. Labels on x axis describe the test conditions, numbers represent the number of rounds of protein handling followed by sanitisation. Top panel shows all results and bottom panel focuses on test and negative controls. Points represent the average of 8 tips and error bards represent standard deviation.

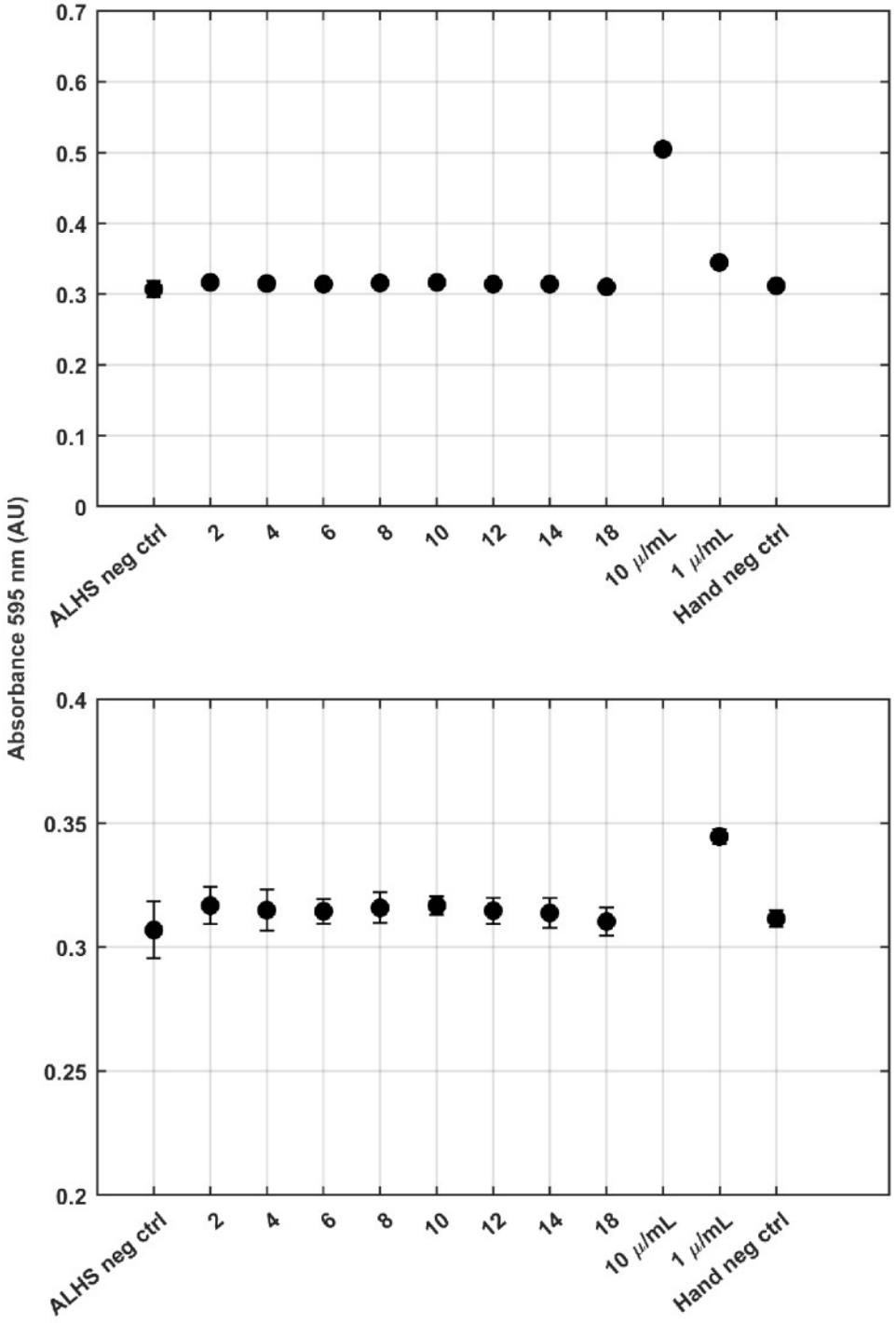

Figure 4-3 Micro-Bradford assay results for sanitisation of Ova. Labels on x axis describe the test conditions, numbers represent the number of rounds of protein handling followed by sanitisation. Top panel shows all results and bottom panel focuses on test and negative controls. Points represent the average of 8 tips and error bards represent standard deviation.

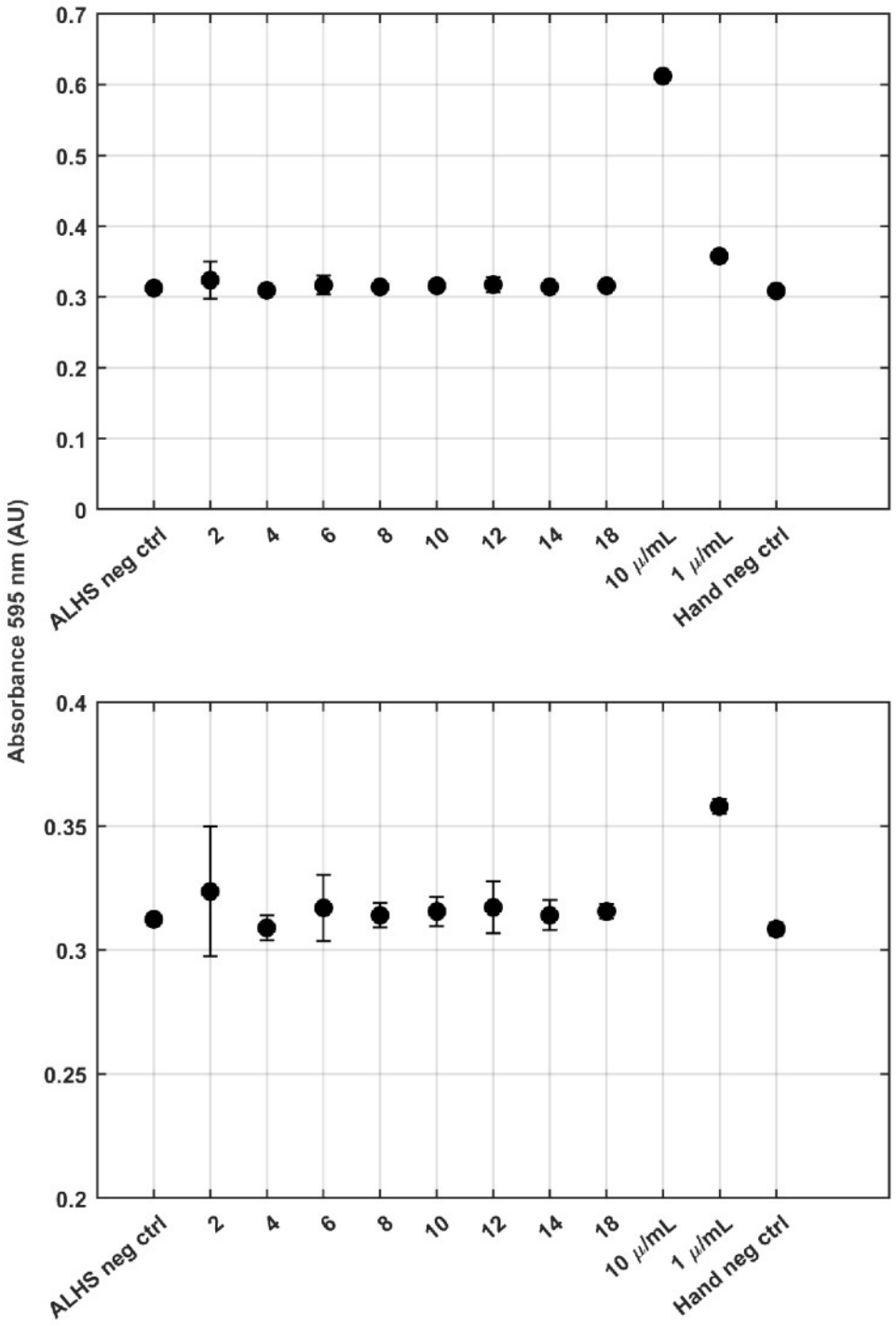

Figure 4-4 Micro-Bradford assay results for sanitisation of BSA. Labels on x axis describe the test conditions, numbers represent the number of rounds of protein handling followed by sanitisation. Top panel shows all results and bottom panel focuses on test and negative controls. Points represent the average of 8 tips and error bards represent standard deviation.

	Hb	*Ova*	*BSA*
P-value	0.00316	0.08528	0.22544

Table 4-1 One-way ANOVA at 0.05 p-value results of micro-Bradford assay comparing 9 test and 2 negative controls. 8 replicates were run, one for each tip.

Carry-over from one protein solution to another was tested by handling 900 µL of protein containing solution twice before sanitisation, 150 µL of 5 mM Tris was then transferred to a test plate. This was repeated so that 9 test solutions were made representing 18 rounds of protein handling. One negative control of 150 µL was transferred before any protein was handled and another was transferred using disposable tips and a hand help multichannel. Two positive control were also added by testing the relevant protein at 1 µg/mL and 10 µg/mL. Before assaying all the samples using a micro-Bradford assay which was chosen for its high sensitivity.

One-way analysis of variants (ANOVA) test was used to statistically test whether there is any difference between the means of two groups: the negative controls and sanitisation test group, the resulting p-values are displayed in Table 4-1. The negative hypothesis tested is that the averages of the Bradford assay test results are the same. The p-values of both Ova and BSA tests suggest the null hypothesis should be accepted, meaning the sanitisation is effective. The null hypothesis must be rejected for Hb, meaning the columns are not the same and sanitisation may not have been effective. This view is supported by Figure 4-2, because the averages of negative controls and test samples look quite different and error bars mostly do not overlap. Visual assessments of Figure 4-3 and Figure 4-4 are is more challenging with some overlapping error bars and some averages appearing similar but others disagreeing. However, comparison of test results with the positive control in Figure 4-2 show that although there may be a small amount of carry-over after sanitisation it is negligible at less than 1 µg/mL, this implies the sanitisation method is fit for purpose. These test results highlight the need to run both positive and negative

controls when doing sanitisation validation as it may not be practically possible to eliminate carry-over entirely but only to reduce it to negligible levels.

To check for carry-over of NaOH from the sanitisation solution to the next liquid handling step the ALHS aspirated and dispensed the same stock of UHQ water 18 times after the sanitisation procedure. Results are displayed in Table 4-2. There was no increase in pH after NaOH sanitisation which implies the final wash step is effective. A pH 7 control solution was run alongside the test samples to ascertain the repeatability of the pH probe which was between 0.03 and 0.10 pH units. Any changes in pH greater than that after washing the tips of NaOH would have warranted greater investigation.

5 + 5 slow	*pH 7 check*	*Ova*	*pH 7 check*	*Hb*	*pH 7 check*	*BSA*
Before	6.95	7.21	7.02	6.51	6.95	7.48
After	7.05	7.10	7.05	6.45	7.00	7.38

Table 4-2 pH testing for NaOH carry-over after sanitisation testing of proteins. pH of water was tested before and after sanitisation alongside a pH 7 control.

4.3 Effect of phase ratio choice on single component isotherm data generation

Experimental parameters such as phase ratio *F* (defined in equation 3-6) are important factors in determining data quality (precision and accuracy defined in equations 4-1 and 4-2) in isotherm studies using slurry plates (Lacki & Brekkan, 2011). Phase ratio can be altered by changing either the volume of liquid or the volume of chromatography resin being contacted. Predictor plates from GE^{TM} Healthcare which are pre-dispensed slurry plates containing a scheduled volume of resin are available for Capto^{TM} Q at resin volumes of 2 µL, 20 µL or 50 µL. The manufacturer's instructions stipulate that 300 µL of liquid should be added to the well. However, this volume can either be reduced (so long as there is enough sample available after incubation for analysis) or increased to a maximum of approximately 500 µL (which allows a sufficient head space after sealing for proper mixing). Moreover, when altering the liquid volume the vigour of agitation must be checked to ensure that the resin is being

fully suspended in liquid and thus mass transfer is not being affected by improper mixing (Bensch et al., 2005).

$$\text{Accuracy} = \frac{\text{Average observed} - \text{Nominal}}{\text{Nominal}} . 100 \qquad \textit{4-1}$$

$$\text{Precision} = \frac{\text{Standard deviation}}{\text{Average observed}} . 100 \qquad \textit{4-2}$$

In order to assess the effect of phase ratio, isotherms were studied at both strongly interacting conditions and weakly interacting conditions. Strong isotherms were generated under conditions were pHs far away from the pI of the protein with a low NaCl level. Weak isotherms were also generated at pHs far from the pI of the protein but in the presence of high concentrations of NaCl. An experimental data point in an SCI consists of a liquid equilibrium concentration expressed along the x-axis and an adsorbed concentration expressed on the y-axis. Traditionally isotherms were quantified by knowing the starting concentration (C_0) of solute and measuring the liquid concentration at equilibrium (C_{eq}), the adsorbed concentration (q) was then calculated via a mass balance as shown in equation 3-3 (Graham L. Skidmore et al., 1990). An additional approach has been explored in this work were the adsorbed protein is eluted and quantified to get a second measure of adsorbed concentration (q^*), the equilibrium concentration can also be quantified in an alternative manner via a mass balance calculation using q^* and C_0 in 3-9. All the isotherms discussed in this chapter are SCIs and so were quantified using spectrophotometry at 280 nm in UV plates in conjunction with a standard curve for the relevant protein.

4.3.1 Calculation of adsorbed protein concentration using measurements taken at equilibrium and after elution

The notation used in the below equations is generally in line with both historic and more recent publications (Faraji et al., 2015; Fritz & Schluender, 1974). Some of the notation referring to mass is new in this area of study but this is the first time quantification of isotherms via elution has been explored (extensive literature search

yielded no publications on this) and so new notation is required. When calculating adsorbed protein concentrations of protein the liquid carried over from one cycle to the next must be considered as both the chromatography resin (V_{resin} in µL) and filter are associated with a some liquid (V_r in µL, equation 3-3). The starting concentration (C_{start} in mg/mL) must be corrected in equation 3-4 to ascertain C_0 in mg/mL once the loaded volume (V_L in µL) has been added to the well. After V_L has been incubated and equilibrium achieved with the resin the flowthrough fractions are collected and the adsorbed protein concentration (q in mg/mL resin) is determined via equation 3-5. Here, q is estimated employing the concentration of the corrected feed solutions (C_o), the protein concentration in the flowthrough (i.e. at equilibrium, C_{eq} in mg/mL) and the phase ratio (F in equation 3-6). In historic publications phase ratio has been described as φ which is the reciprocal of F (Conder & Young, 1979; Graham L. Skidmore et al., 1990), F has been used in more recent publications, particularly in reference to work using slurry plates (Chhatre & Titchener-Hooker, 2009).The analysis of the collected wash and elution factions returned a second measure of adsorbed protein concentrations (q^* mg/mL of resin) via equation 3-7. Here, the carried over protein mass (m_r in µg) in each fraction is calculated by multiplying the concentration (C in mg/mL) by V_r each time the filter plates are evacuated (equation 3-8). Hence, the i^{th} fraction will carry-over a given amount of protein to fraction *i+1.* The protein mass in all wash (m_w in µg) and elution (m_e in µg) is calculated before subtracting the mass retained from the protein equilibration step (m_{req} in µg), the mass carried over from the wash steps (m_{rw} in µg) and the mass carried over from the elution steps (m_{re} in µg). The corrected elution mass is then divided by V_{resin}. C_{eq}^* (mg/mL) is the equilibrium concentration calculated via mass balance from q^* as described in equation 3-9.

$$V_r = 0.6 \times V_{resin} + 6 \quad \textit{3-3}$$

$$C_0 = \frac{C_{start} \times V_L}{V_L + V_r} \quad \textit{3-4}$$

$$q = (C_0 - C_{eq}) \times F \quad \textit{3-5}$$

$$F = \frac{(V_L + V_r)}{V_{resin}} \quad \textit{3-6}$$

$$q^* = \frac{(\sum_{i=1}^{n} m_{w,i} - m_{rw,i} + \sum_{i=1}^{n} m_{e,i} - \sum_{i=1}^{n-1} m_{re,i} - m_{req})}{V_{resin}} \quad \textit{3-7}$$

$$m_{r,i} = V_r \times C_i \quad \textit{3-8}$$

$${C_{eq}}^* = C_0 - \frac{q^*}{F} \quad \textit{3-9}$$

4.3.2 Phase ratio comparison in strongly interacting conditions

In order to corroborate the data generated using slurry plates, a larger-scale batch experiment was performed. 10 mL of resin was used prepared and used in a kinetic uptake experiment monitoring protein uptake over time at 50 mM Tris pH 9 with 50 mM NaCl. As a comparison plate data was generated in the same buffer conditions using 20 µL at 3 phase ratios by varying the volume of protein loaded. The results are displayed in Figure 4-5. The results from the larger-scale batch adsorption and micro-scale plate experiments agree very well. The plate isotherms where incubated for 90 minutes and the uptake experiment was run at highly overloading conditions (300 mg/mL resin) the 90 min uptake data point can be directly compared with any points on the plateau of the isotherm. The saturation capacity agrees well across formats despite the fact that the volume of resin used in these experiments varies by a factor of 500. The uptake data in Figure 4-5 suggests that 90 min is a good period for incubation: A short incubation is of course preferred as it reduces experimental time and increases throughput which is important as multicomponent studies require multiple plates to be run. However, the incubation period needs to be long enough

so that a quasi-equilibrium state can be achieved whilst reducing experiment time. Figure 4-5 shows that the rate of uptake has significantly slowed and is has reached plateau by 90 minutes.

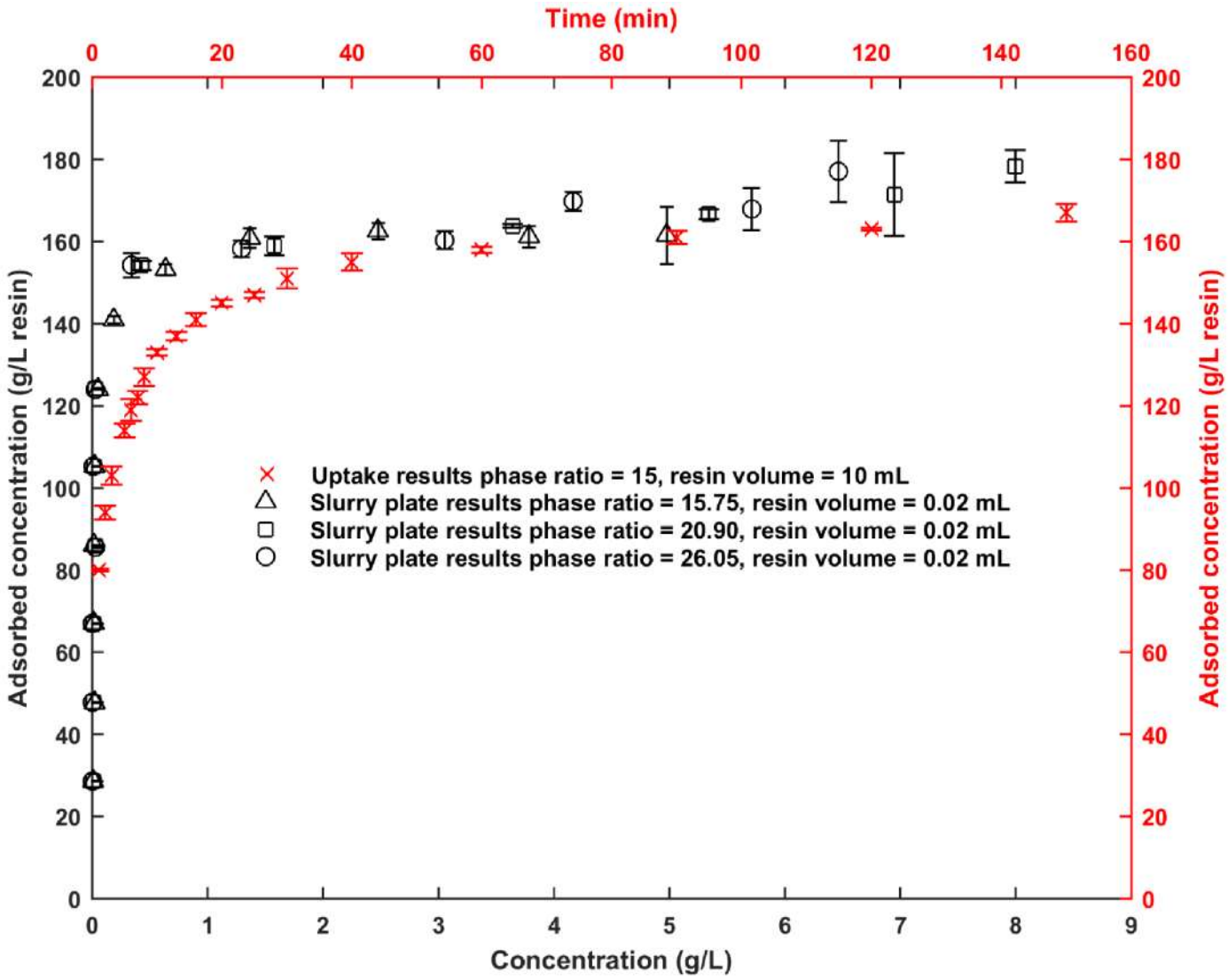

Figure 4-5 Comparison of equilibrium data generated on a slurry plate at 3 phase ratios with kinetic uptake data generated at beaker scale using BSA at pH 9 50 mM Tris and 50 mM NaCl. Plate data was generated using 20 µL of resin and liquid volume was varied in order to achieve different phase ratios. Proteins were incubated for 90 minutes on the plate. Uptake data was generated using 10 mL of resin as described in the legend, protein was overloaded for uptake at 300 mg/mL resin. Both primary and secondary y-axis are the same, but the primary and secondary x-axis are different for the isotherm and uptake data. The colour of the data matches the colour of the axis which the data should be read on.

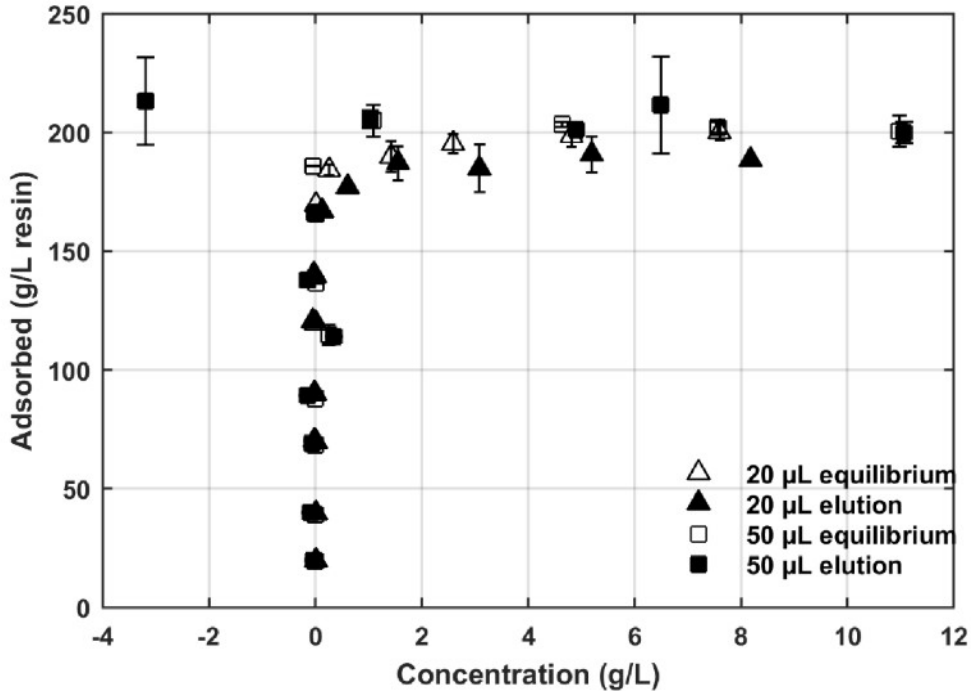

Figure 4-6 BSA isotherms in strongly interacting condition at different phase ratios. Conditions tested were 50 mM Tris pH 9 0 mM NaCl, phase ratio was altered by changing the volume of resin as described in the legend, liquid volume was maintained at 400 µL. Equilibrium data is plotted as C_{eq} vs q, elution data is plotted as C_{eq}^{*}vs q^{*}.

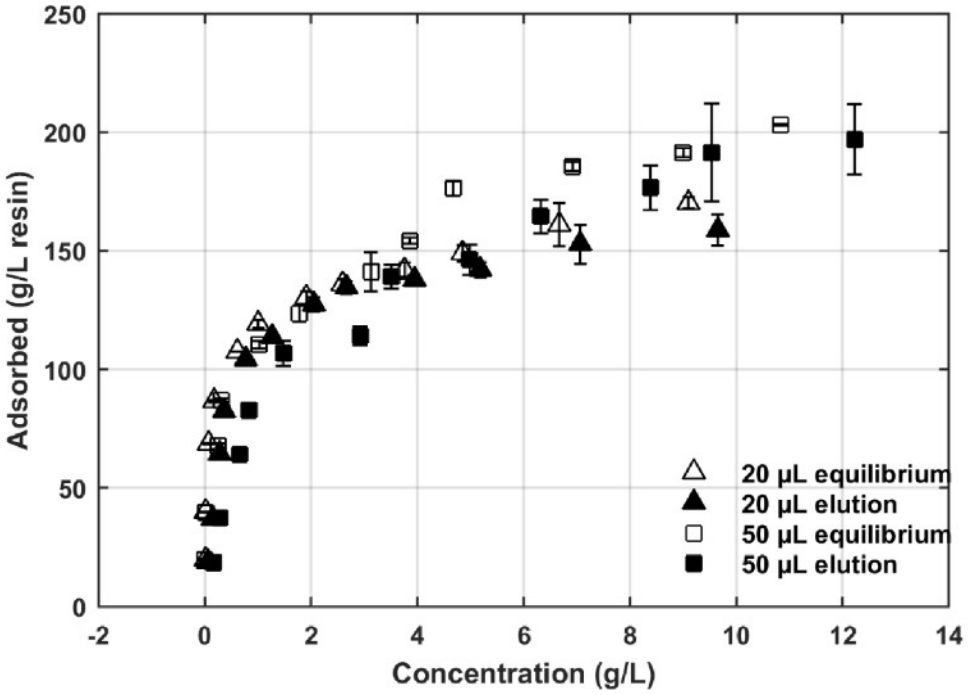

Figure 4-7 Ova isotherms in strongly interacting condition at different phase ratios. Conditions tested were 50 mM Tris pH 9 0 mM NaCl, phase ratio was altered by changing the volume of resin as described in the legend, liquid volume was maintained at 400 µL. Equilibrium data is plotted as C_{eq} vs q, elution data is plotted as C_{eq}^{*}vs q^{*}.

As it has been shown above that 20 µL of resin is suitable for studying strongly interacting isotherms and that it is robust to changes in incubated liquid volume the effect of increasing the resin volume for 20 to 50 µL was studied for strongly interacting isotherms.

Strongly interacting isotherms were generated for BSA and Ova at pH 9 with 0 mM NaCl. The same isotherms were compared whilst changing the phase ratio with resin volume alternating between 20 and 50 µL with liquid volume remaining constant at 400 µL, isotherms were calculated at equilibrium using C_{eq} and q as well as via elution C_{eq}^{*} and q^{*}. Results for BSA and Ova are displayed in Figure 4-6 and Figure 4-7 respectively.

There is one outlier in Figure 4-6 which gives a negative concentration in solution. In order to generate measures for both adsorbed and equilibrium concentrations which are independent as possible, equilibrium concentrations which are associated with elution data points are calculated via mass balance (C_{eq}^{*}) from the adsorbed concentration instead of direct observation as in the case of equilibrium data points. For the outlier the adsorbed concentration has been overestimated due to a random error in assaying one elution fraction resulting in a negative equilibrium concentration. Isotherms are is in good agreement using both resin volumes and across equilibrium and elution meaning that either 50 and 20 µL could be used for isotherm studies looking at strongly interacting conditions. It is preferable to use the smallest resin volume possible as it saves material. Additionally, in order to saturate 50 µL of Capto^{TM} Q very high concentrations of protein are required (50 mg/mL), as mass cannot be further increased by loading a larger volume due to the size of the well in the 96 well plate, the effect of increasing protein stock concentration will be discussed later. An alternative, which was explored for Figure 4-6 and Figure 4-7, multiple rounds of loading incubation are required in order to achieve high loads. This is disadvantageous from a time/throughput perspective as 2 rounds of 90 min incubation are required, as previously mentioned throughput is an important factor for multicomponent studies which require many plates. Multiple loads also add further opportunities for experimental error to occur. Errors when loading a volume

into plate cannot be circumvented by performing alternative dilutions or reanalysing samples. An event when preparing the feed solutions may be noticed and flagged as the feed plate is always analysed, errors in dilution and analysis can be subject to repeat dilution and/or analyses and alleviated in that way. There is no such opportunity of there is an error in loading the plate, as such opportunities for these errors to occur should be minimised. Therefore it is better to use 20 µL of resin rather than 50 µL in order to increase repeatability and minimise protein usage and facilitate accurate liquid handling.

4.3.3 Phase ratio comparison for moderately interacting isotherms

The strength of protein interaction can be thought of as strong, moderate or weak. Strongly interacting isotherms are associated with rectangular shapes with increases in adsorbed concentration and very little increase in equilibrium liquid concentration until the plateau is reached, an example if which is displayed in Figure 4-6. Strongly interacting isotherms are observed in ion exchange chromatography at pHs further away from pI and more weakly interacting isotherms are found either at a pH close to the pI of the protein or at a pH far away from the pI but in the presence of higher levels of NaCl.

In column separations, if there were no change in the salt level strong isotherms would reflect irreversible binding, or elution so slow it would not be observed during a chromatographic run. For elution to occur in a sensible time frame the isotherm must be supressed by the addition of a modifier such as NaCl which reduces the initial slope of the isotherm and causes elution. The retention factor *k'* which is defined in equation 4-3 and is an indicator of this interaction strength as it is a measure of the isotherm slope in its linear region. It has been suggested that *k'* values greater than 10^2 represent strong isotherms with irreversible binding, *k'* values less than 1 represent weak binding, anything in between these values is considered moderate. The reason for this extremely strong interaction is the multi-pointed nature of adsorption with multiple counter ion sites on the protein interacting with the ligands giving a very strong aggregate interaction, salt counter-ions interrupt this

interaction eventually causing the isotherm to become weak (Velayudhan & Horváth, 1988).

Imagine the separation of a binary mixture of proteins which are strongly interacting with the sorbent at low modulator concentrations, these proteins are about to be eluted by a linearly increasing gradient of NaCl. As the resolution of these compounds is dependent on the ratio of their *k'* values the very large values *k'* ($> 10^2$) for both proteins means effectively no travel down the column occurs, therefore there is no meaningful separation. As the NaCl concentration increases the isotherm slopes are reduced along with their *k'*, it is now the ratio of these *k'* values which defines their resolution. If the NaCl concentration were to increase further both isotherms would have low *k'* values (< 1), again meaning no further separation would occur. As such the critical isotherms to capture are the moderate isotherms as they determine resolution (Velayudhan & Horváth, 1988).

Observation of equation 3-5, which is the traditional method of obtaining adsorbed concentration, shows the equilibrium concentration is subtracted from the starting concentration and then multiplied by the phase ratio. In more weakly interacting isotherms the difference between the start and equilibrium concentrations is less than in more strongly interacting and as both the values are subject to random error the propensity for noise is greater for more weakly interacting isotherms. Analysis of Figure 4-8 shows the isotherm has a *k'* of 15 meaning it is a moderate isotherm (Velayudhan & Horváth, 1988).

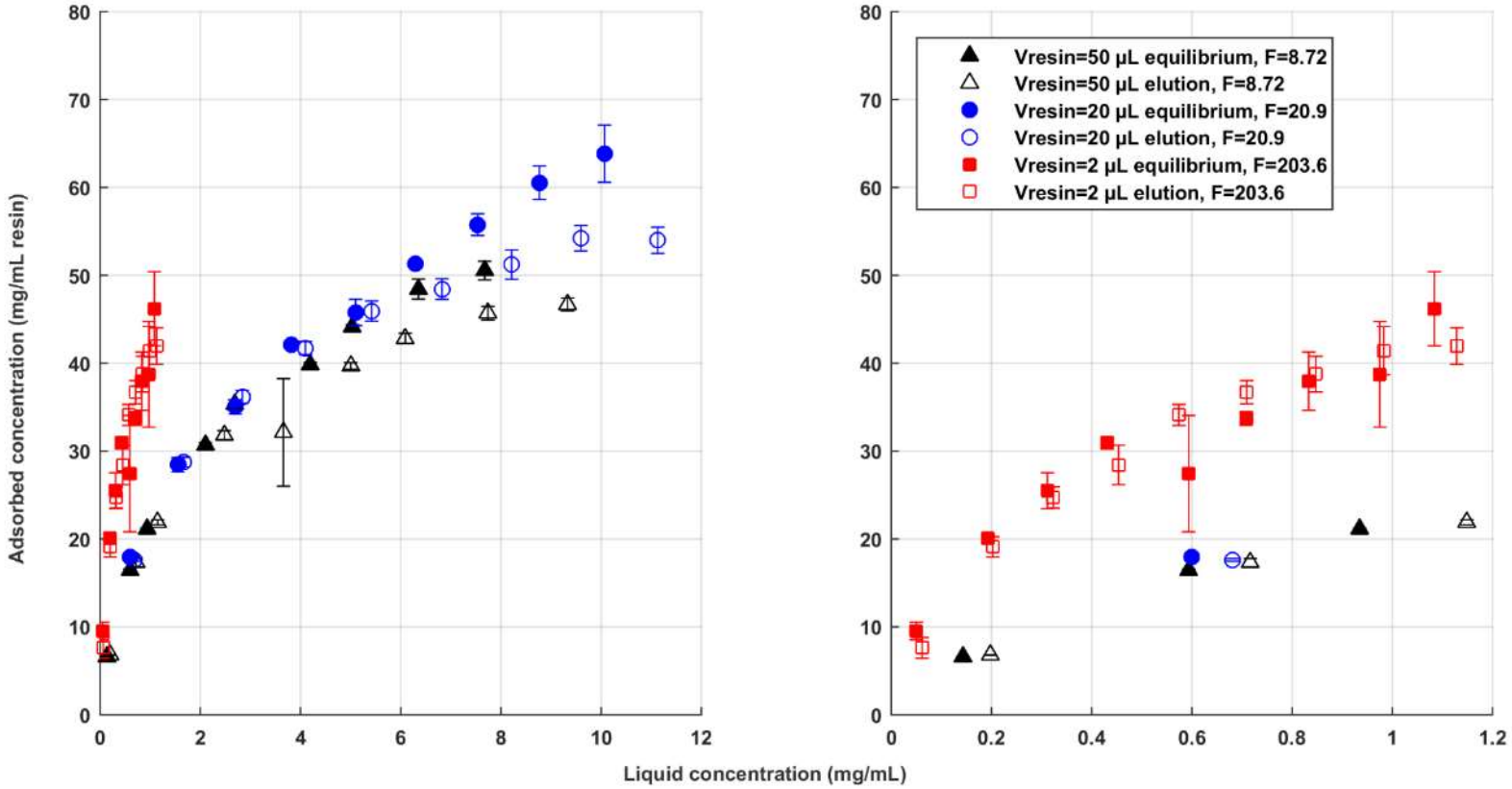

Figure 4-8 Bovine serum albumin (BSA) isotherms in 200 mM NaCl pH 9 50 mM Tris at 3 phase ratios. Isotherms were calculated from both equilibrium and elution measurements. The liquid volume was kept constant at 400 µL, resin volume (V_{resin}) and phase ratios (F) are described in the legend. Left hand panel compares results from all phase ratios, right hand panel displays the same data but with the x axis expanded to focus on the 2 µL data. Error bars represent standard deviation of experimental replicates.

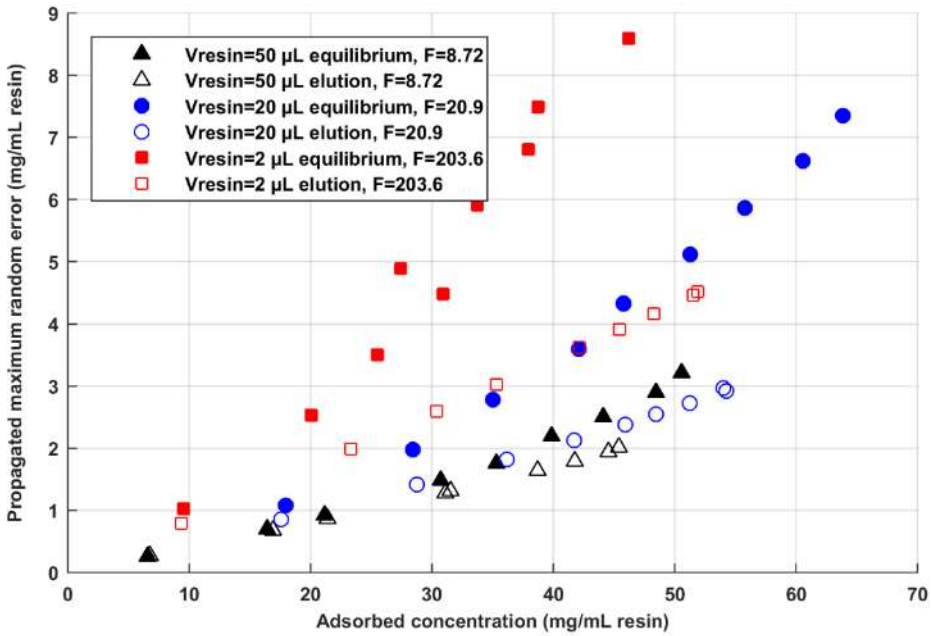

Figure 4-9 Propagated maximum precision error in adsorbed concentration at 3 phase ratios. Errors were propagated based on isotherms displayed in Figure 4-8 which is BSA pH 9 200 mM NaCl using 2, 20 and 50 µL of resin and keeping loaded protein volume constant at 400 µL.

$$k' = \frac{q}{C_{eq}}/F \qquad 4\text{-}3$$

Moderately interacting isotherms for BSA pH 9 50 mM Tris 200 mM NaCl were generated at 3 phase ratios. Phase ratio was varied by changing the resin volume between 2, 20 and 50 µL but keeping the volume of protein constant at 400 µL, data was generated both at equilibrium and by via elution, results are displayed in Figure 4-8. Figure 4-8 shows that data generated using both 20 and 50 µL of resin agrees well across the phase ratios. Additionally the *q** agrees well with the *q* data. Using 2 µL of resin resulted in a systematic over-estimation in the adsorbed concentration for the isotherm. Interestingly the *q* and *q** data is internally consistent within phase ratios despite the disagreement between phase ratios. It may be that there are some non-specific interactions between the protein, the plate, filter or plate sealer; such interactions have previously been observed (Bhambure & Rathore, 2013). Such interactions may then be eliminated during elution in high NaCl meaning that the

equilibrium and elution data agree within the 2 µL data set. These non-specific interactions also occur when using smaller phase ratios, but the structure of equations 3-5 and 3-7 means that any systematic error is exaggerated with smaller resin volumes and larger phase ratios. As previously mentioned calculating adsorbed concentration in moderately interacting isotherms is error prone due to the subtraction of two which are close to another and are subject to error. That difference/error is then multiplied by the phase ratio meaning large phase ratios amplify this error more than small ones. This would be the case both for both random error which affects precision and for systematic error in the case of non-specific interactions, which affects accuracy.

Random errors have been propagated for data displayed in Figure 4-8 and is presented in Figure 4-9, the method of propagation is discussed in Chapter 3. In short the maximum error in q and q^* is estimated using the random error observed for each term of equations 3-5 and 3-7 and propagating them using equations in Table 3-1.

Figure 4-9 shows the propagated error in adsorbed concentration at the 3 phase ratios calculated for q and q^*. Considering only q, Figure 4-9 shows a positive correlation between the noise level and the phase ratio, F. This was also observed in Figure 4-8 and has also been discussed elsewhere (Lacki & Brekkan, 2011). This correlation is due to the multiplication by F in Equation 3-5. When high levels of noise are expected, such as in moderate or weakly binding conditions, the analysis of the elution fractions can produce isotherms with reduced variability. Propagated error of adsorbed concentration via elution is less than when compared from equilibrium at the same phase ratio when isotherms are produced using the methodology described here; when calculated from equilibrium the calculation involves subtracting 2 values which are close to one another (due to weaker adsorption) and subject to variability, whereas elution involves addition of small values which are also subject to variability. Propagated error in adsorbed concentration calculated via elution picks up dramatically when resin volume decreases to 2 µL due to increased resin volume variability.

Although variability of adsorbed concentration is rarely reported numerically, precision in *q* was reported as mainly between 1.6% and 2.1% (a. Osberghaus et al., 2012). When comparing variability with our own data using an isotherm of similar shape and adsorption strength (BSA pH 9 100 mM NaCl) we find observed experimental precision varying between 0.0% - 3.6% in *q* and 0.6% - 3.0% in *q**.

Larger resin volumes associated with reduced variability and improved accuracy require increased protein stock concentrations in order to explore binding capacity. The effect of these increased protein stock concentrations on liquid handling is explored below.

From surveying the data so far it is clear that larger resin volumes are preferable as they are associated with less systematic and random error. However, there are drawbacks to using the largest resin volume (50 µL) as it requires additional protein to saturate the resin. Either a much increased protein stock concentration must be used or increased incubation periods. 20 µL offers major improvements in systematic and random error and does not require such excessive protein stock concentrations or increased incubation periods; therefore it is the natural choice for future studies. However, increasing the volume of resin from 2 µL to 20 µL does require an increased protein stock concentration from 2 to 20 mg/mL. This increase in stock concentration can affect liquid handling performance of the ALHS as discussed below.

4.4 Liquid handling assessment

As previously mentioned increased resin volumes require increased protein concentrations in order to saturate. A gravimetric liquid handling assessment of the ALHS was performed for the solutions required to perform isotherm determination experiments using either 2 or 20 µL of resin. Results of the assessment are displayed on Table 4-3. The accuracy and precision for all solutions with ≤ 2 mg/mL protein content showed precision and accuracy within 1% (1 significant figure) at volume ≥ 50 µL. However, when using the same default liquid handling class both BSA and Con at 20 mg/mL the liquid handling accuracy exceeds 1% at volumes between 170 and

20 µL. In order to alleviate this issue separate custom liquid handling classes were setup from both BSA and Con at 20 mg/mL as described in the literature (Treier et al., 2012). Liquid handling performance for these solutions after liquid handling calibration is also displayed in Table 4-3. After solution specific-calibration for BSA and Con at 20 mg/mL, the accuracy closes to ≤ 1% at volumes of 50 µL or more.

In conclusion, increasing protein stock concentration leads to decreases in liquid handling accuracy. This issue was resolved by setting up solution-specific liquid handling classes to avoid systematic errors in isotherm determination. Additionally the liquid handling assessment shows that handling volumes less than 50 µL should be avoided as accuracy and precision drops off below that volume.

		Accuracy (%)				*Precision (%)*			
		20 µL	***50 µL***	***170 µL***	***700 µL***	***20 µL***	***50 µL***	***170 µL***	***700 µL***
Calibrated liquid handling class	***BSA 20 mg/mL***	0.66	-0.35	-0.25	-0.14	0.98	0.5	0.2	0.14
	Con 20 mg/mL	-1.92	-0.33	-0.39	-0.29	3.95	0.55	0.41	0.09
Standard liquid handling class	***Water***	-1.52	0.85	0.06	-0.11	3.43	0.81	0.17	0.12
	BSA 20 mg/mL	-5.98	-3.33	-2.8	-0.14	1.67	0.67	0.37	0.3
	Con 20 mg/mL	-6.8	-3.44	-1.67	-0.17	3.39	1.14	0.16	0.35
	Ova 20 mg/mL	-1.84	-0.07	-0.83	-0.2	0.77	0.3	0.11	0.05
	BSA 2 mg/mL	-3.95	-0.93	-0.67	-0.1	2.45	0.94	0.46	0.33
	Con 2 mg/mL	-5.06	-1.22	-0.79	-0.12	2.17	0.99	0.35	0.3
	Ova 2 mg/mL	-4.13	-0.64	-0.4	-0.15	3.71	0.59	0.47	0.32
	1 M Tris	-0.79	-0.19	-1.22	-0.28	0.34	0.42	0.16	0.05
	2 M NaCl	-2.16	-0.16	-0.92	-0.48	0.78	0.24	0.09	0.04
	5 mM Tris	-0.89	0.83	-0.2	-0.41	0.22	0.11	0.05	0.01

Table 4-3 Liquid handling performance of ALHS for solutions requited during isotherm generation. Accuracy and precision (defined in equations 4-1 and 4-2) of solutions before and after solution specific calibration.

4.5 Conclusions on critical factors affecting isotherm data generation

Once fixed tips were selected as the best choice for the application of isotherm data generation, sanitisation procedures were investigated to ensure there was negligible protein carry-over between liquid handing steps, and also to ensure NaOH carry-over was not observed.

Initially it was shown that the predictor plate experimental setup using 20 µL agrees well with experiments done at a much larger scale using 10 mL of resin. It was then shown that either 20 or 50 µL isotherms agree well for strongly interacting isotherms. Once the HT slurry plate method had been shown to work effectively for strongly interacting isotherms which are comparatively less prone to error than weakly interacting isotherms, moderately interacting isotherms were then investigated. It was then shown that the 2 µL of resin in weakly interacting conditions results in much higher propensity for error than when using larger resin volumes and that their use resulted in systematic over estimation in adsorbed concentration meaning only 20 or 50 µL of resin should be used for studies. It was also discussed that whilst 50 µL could be used for studies the larger volume of resin is associated with either increased protein stock concentrations or decreased throughput for experimental activities which is a consideration when studying multicomponent isotherms requiring many plates to be run. Additionally 50 µL of resin increases use of expensive proteinaceous material.

Finally liquid handling capability of the ALHS was assessed for solutions required for isotherm studies. It was shown that liquid handling performance begins to drop off when using high stock concentrations required to saturate 20 µL of resin and also that handling volumes below 50 µL should be avoided. Liquid handling performance was restored by calibrating specific liquid handling classes for the errant solutions.

Studies presented in the chapter suggest 20 µL should be used for isotherm studies as the use of 2 µL results in systematic errors when studying moderately

interacting isotherms. After a careful review of the data it was decided that 20 µL of resin will be used for studies and solution specific calibrated liquid handling classes will be used when required.

4.6 Single component column and isotherm data

Three model proteins were taken forward for column runs and isotherm studies. The properties of these proteins and reasoning for their choice will be discussed before the presentation of column and isotherm data.

4.6.1 Model protein properties and behaviour on column

The proteins whose properties are discussed in Table 2-2 were chosen for these studies because firstly because their isoelectric points (pI) are relatively close, although pI is not an absolute predictor retention on an anion exchanger it does give some indication. Secondly the proteins are all relatively large with the smallest protein being Ova at 45 kDa. Previous studies have focussed on the typical model proteins which tend to be relatively small such as myoglobin (17 kDa), lysozyme (14 kDa), ribonuclease a (14 kDa) (Chen, Hu, & Wang, 2006; Katiyar, Ji, Smirniotis, & Pinto, 2005; A. Z. Osberghaus, 2012; Roth et al., 1996). This is somewhat smaller than a typical bio-therapeutic such as a monoclonal antibody which is approximately 150 kDa. Although none of the proteins are as large as 150 kDa it is hoped that the larger size of these model proteins will at least reflect some of the increased complexity of interaction associated with larger proteins. Finally these proteins are readily available as lyophilised powder from suppliers, highly soluble and relatively cheap.

Preparative chromatography runs of the model proteins with a relatively steep NaCl gradient for elution are displayed in Figure 4-10, proteins were loaded at low level, well below saturation in the linear region of the isotherm. Con shows a high absorbance than the BSA and Ova only because its absorbance in AU/mg/mL is greater than BSA and Ova which have similar slopes in their 280 nm calibration curves. Chromatogram shows that the proteins are indeed quite similar in retention on the strong anion exchanger at pH 9, peak maxima at within 2 CVs of the next protein: 10.4, 12.0 and 13.0 CV for Con, Ova and BSA respectively. Moreover, the percent overlap of each species with one another shown in Table 4-4 suggests this is rather a competitive system, adjacently eluting proteins have overlaps of at least 10% and there is also overlap between the most at least retained species. It should be noted that whereas the maxima of the eluted peak is dependant only on the isotherm or equilibrium interaction the peak overlap is dependent on the retention but also the peak width, the peak width or band spreading is dependent upon mass transport kinetics during elution which is independent of the isotherm at these low loads in the linear region of the isotherm.

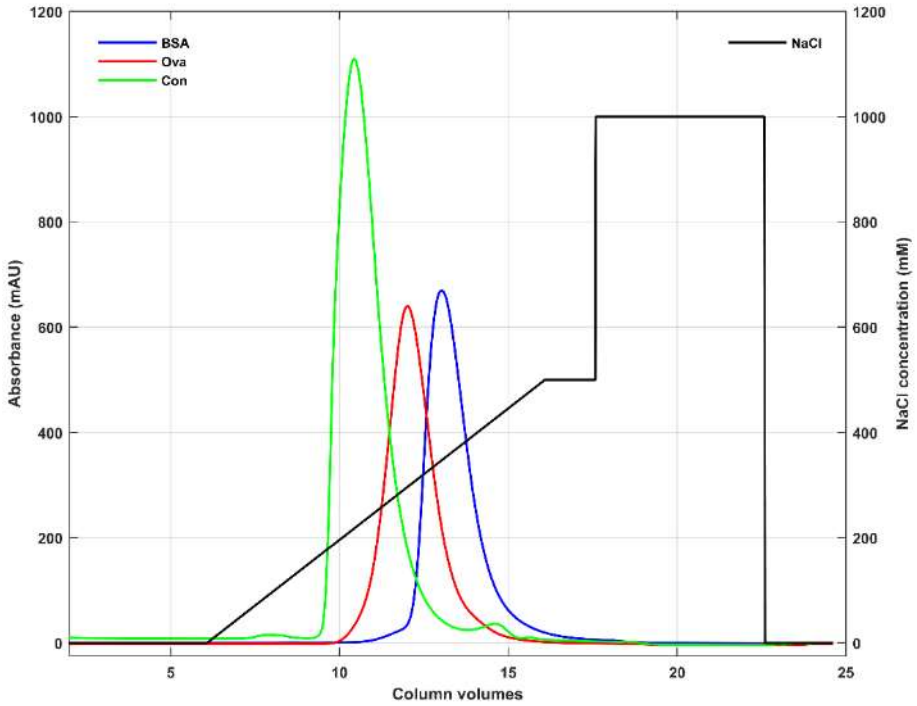

Figure 4-10 Chromatograms from 3 separate overlayed AKTA runs of BSA, Ova and Con. HiScreen™ column from GE™ Healthcare using 4.7 mL of Capto™ Q 10 cm column at run pH 9 50 mM Tris with a NaCl gradient of 50 mM/CV starting at 0 mM with a linear velocity of 260 cm/hr. The load was 11 mg/mL resin and injection volume was 5 mL.

Overlap identity	*% peak area overlap*
***Con**-Ova*	16%
***Con**-BSA*	6%
***Ova**-Con*	16%
***Ova**-BSA*	27%
***BSA**-Ova*	12%
***BSA**-Con*	4%

Table 4-4 Overlap of proteins with one another. The overlap which the percentage is refers to is written in bold.

4.6.2 Single component isotherms

BSA, Ova and Con isotherms were elucidated using PreDictor^{TM} plates and the ALHS as discussed previously. 20 µL of Capto^{TM} Q was used with a corrected phase ratio of 20.9 for the remaining isotherms presented in this chapter. 20 µL of resin was chosen and liquid specific calibrated liquid handling classes were used for the both BSA and Con at 20 mg/mL as recommended earlier. Adsorbed protein concentration was calculated both via mass balance from equilibrium measurement (q equation 3-5) as well as by assaying eluted fractions (q^* equation 3-7).

All the SCIs were generated in 50 mM Tris at either pH 8 or 9. The NaCl concentrations studied are displayed in Table 4-5 and the isotherms themselves are displayed in Figure 4-11. A comparison of propagated error in the equilibrium concentration as measured directly (C_{eq}) and via elution (C_{eq}^* in equation 3-9) is displayed in Figure 4-12. Assessment of Figure 4-12 shows that at all pHs and NaCl levels there is less propagated precision error associated with direct measurement (C_{eq}) than there is associated with C_{eq}^* (equation 3-9) which is the equilibrium concentration calculated via mass balance from elution data. All SCIs and MCIs from this point on, including Figure 4-11, will plot on the x-axis the equilibrium concentration measured directly (C) and not via mass balance from elution (C_{eq}^*).

BSA isotherms at both pH 8 and 9 are typical of a strongly interacting species with very little BSA present in solution at equilibrium until the saturation capacity is reached. As the NaCl concentration increases to 100 mM at pH 9 or 50 mM at pH 9, the depletion of nearly all BSA in the liquid phase continues until the saturation capacity is reached. At the highest NaCl levels studied the rectangularity of the isotherm decreases significantly as BSA left in solution and adsorbed to the resin at the same time. There is little difference between pH 9 and pH 8 BSA isotherms except that pH 8 has increased NaCl sensitivity.

Ova behaves rather differently to both BSA and Con as it has a less rectangular corner, which will be referred to as the or knee, even in the presence of no NaCl,

additionally the isotherm plateau continues to steadily rise as equilibrium concentration increases. Again the Ova isotherm is very similar at pH 8 and 9 but with increased NaCl sensitivity at pH 8. Ova NaCl sensitivity is similar to that of BSA with similar levels of isotherm suppression.

Con isotherm at pH 9 has a similar behaviour to BSA with low levels of Con left in solution until the saturation capacity is reached. It loses this behaviour at 50 mM NaCl, much earlier than BSA stops showing that behaviour. Con is also significantly more sensitive to increases in NaCl concentration at pH 8 than the other proteins as it is well supressed at 100 mM NaCl as opposed to 200 mM NaCl. Con is also much more sensitive to changes in pH than both BSA and Ova. The isotherm loses its tendency to deplete nearly all the protein left in solution before reaching the plateau at pH 8 in the presence of no NaCl.

		BSA	Ova	Con
NaCl (mM)	*pH 9*	0 - black	0 - black	0 - black
		-	-	25 - red
		100 - red	100 - red	50 - blue
		200 - blue	200 - blue	100 - pink
	pH 8	0 - black	0 - black	0 - black
		50 - red	50 - red	-
		100 - blue	100 - blue	25 - red
		150 - pink	150 - pink	50 - blue

Table 4-5 Concentration of NaCl and pH of 50 mM Tris solutions employed in the measurements of single component isotherms for BSA, Ova and Con. Colours describe plotted single component isotherms shown in Figure 4-11 and Figure 4-12.

The behaviour seen in the SCIs agree well with both the reported pIs of the protein and order of elution during chromatography runs in addition to other published data on the elution behaviour of these proteins (Kopaciewicz et al., 1983). The proteins elute in order of NaCl sensitivity and their sensitivity is in agreement with their pIs, BSA whose pI is furthest away from pH 8 has the lowest sensitivity and Con which has its pI closest to 8 is most sensitive. Con is also the most sensitive pH which is unsurprising as its pI is close to the pHs studied. Although pI has been an

effective predictor of NaCl sensitivity here that is not necessarily always the case. Protein interaction can be the product of local patches of charge which dominate the interaction, these patches can remain despite the average charge of the protein changing as has been discussed elsewhere (Shi, Zhou, & Sun, 2008; Yamamoto & Ishihara, 1999).

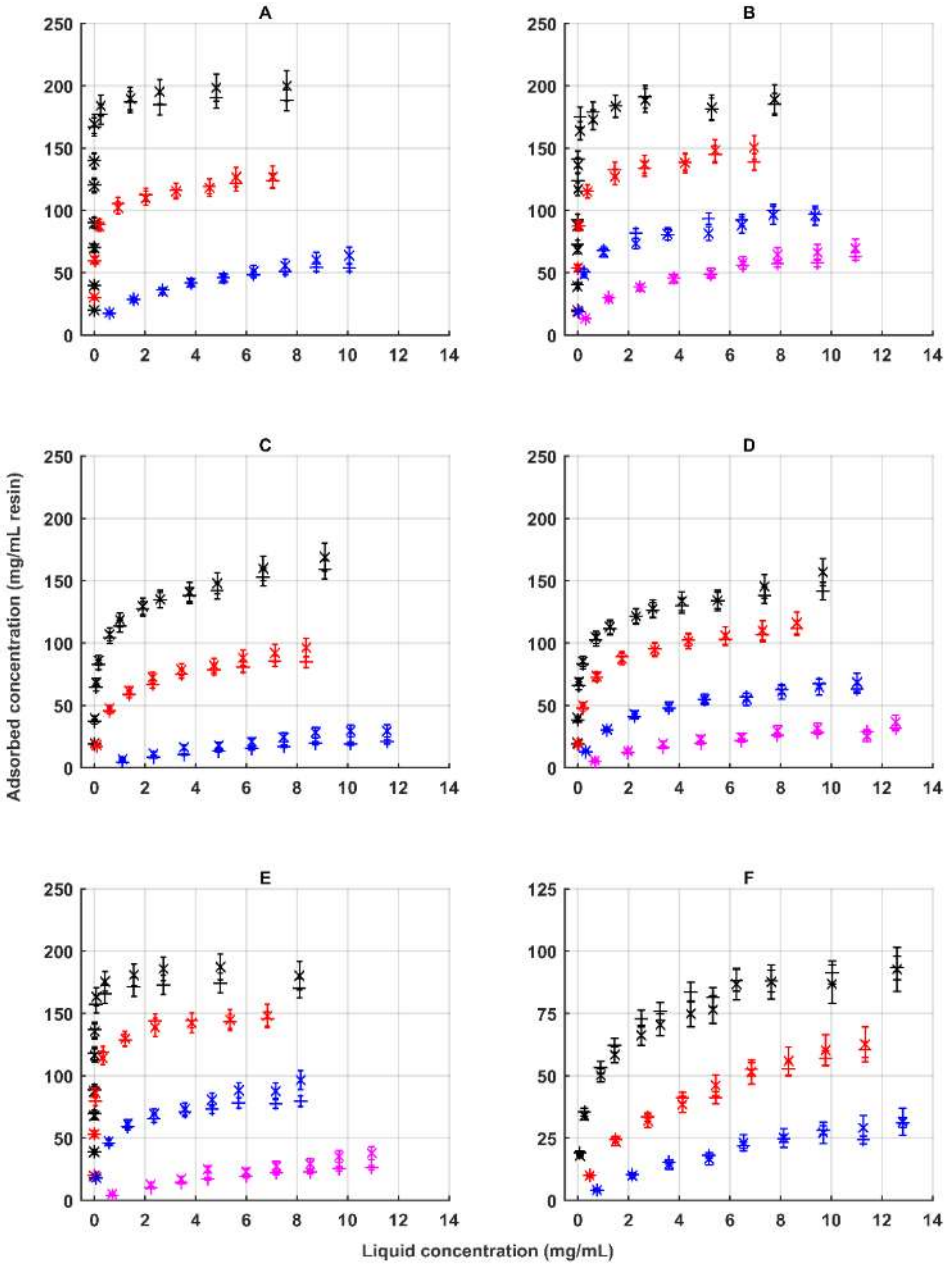

Figure 4-11 Single component isotherms of BSA, Ova and Con at pH 8 and 9 and varying levels of NaCl generated at equilibrium and via elution. Panels A, C & E show isotherm data generated at pH 9 and panels B, D & F at pH 8 50 mM Tris. Panels A & B show BSA isotherms, panels C & D show Ova isotherms and panels E & F show Con data. Cross symbols represent equilibrium data and plus symbols represent elution data, error bars represent maximum propagated error in precision in adsorbed concentration. The concentration at equilibrium displayed in the x axis assayed directly for both equilibrium and elution data. Different colours represent NaCl levels for isotherms displayed and are described in Table 4-5.

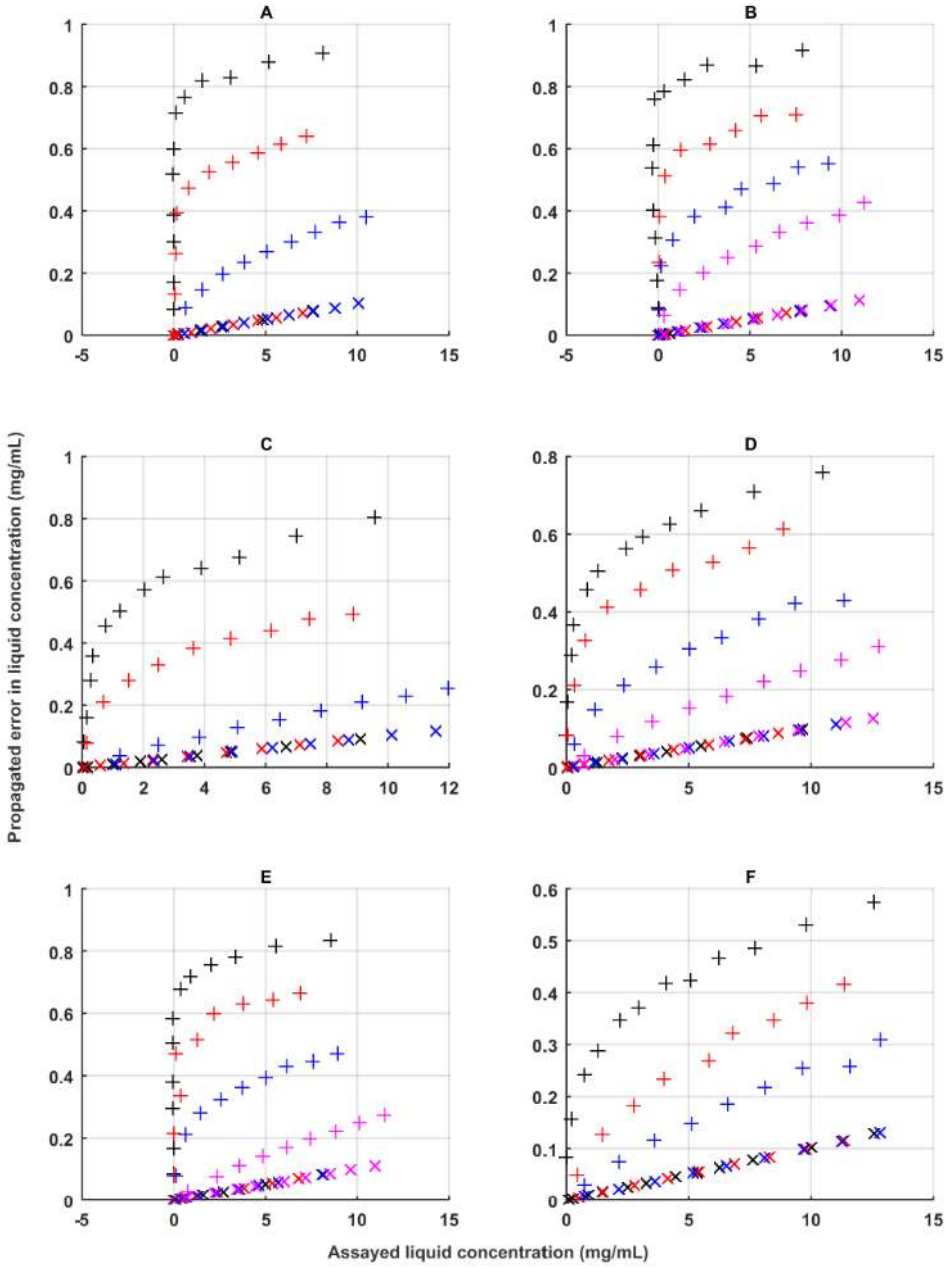

Figure 4-12 Propagated error in liquid concentration assayed directly (C_{eq}) and via mass balance from elution (C_{eq}^{*} in equation 3-9) for BSA, Ova and Con at pH 8 and 9 and varying levels of NaCl generated at equilibrium and via elution. Panels A, C & E show data generated at pH 9 and panels B, D & F at pH 8 50 mM Tris. Panels A & B show BSA isotherms, panels C & D show Ova isotherms and panels E & F show Con data. Cross symbols represent equilibrium data and plus symbols represent elution data. Different colours represent NaCl levels for isotherms displayed and are described in Table 4-5.

4.6.3 Comparison between isotherms calculated at equilibrium and via elution

Agreement between equilibrium measurement and elution measurements is generally very good. Figure 4-11 shows that the error bars which represent maximum propagated precision error overlap for most of the isotherms. Inspection of Figure 4-11 and Figure 4-13 at pH 9 shows there is some tendency for the adsorbed concentration to be underestimated under elution on the plateau when the saturation capacity is high. There is also some under estimation for the more weakly interacting isotherms at pH 9. There was some precipitation observed for the most strongly interacting isotherms in the 1st elution fraction at pH 9, this is may have been because at pH 9 the 1st elution cycle was at 1 M NaCl. To counter this the pH 8 isotherms the first elution cycles were carried out at 0.3 M NaCl and the subsequent cycle at 1 M NaCl, no precipitation was observed in pH 8 elution cycles. pH 9 SCI elution data was generated using a wash step using the equilibration buffer to wash any unbound protein, this may have the slight underestimation of protein bound via elution as small underestimations in protein concentrations at the bottom assayable range of the standard curve may add up to affect mass balance closure. For pH 8 the wash steps were eliminated and better agreement between equilibrium and elution is achieved for the estimation of adsorbed concentration.

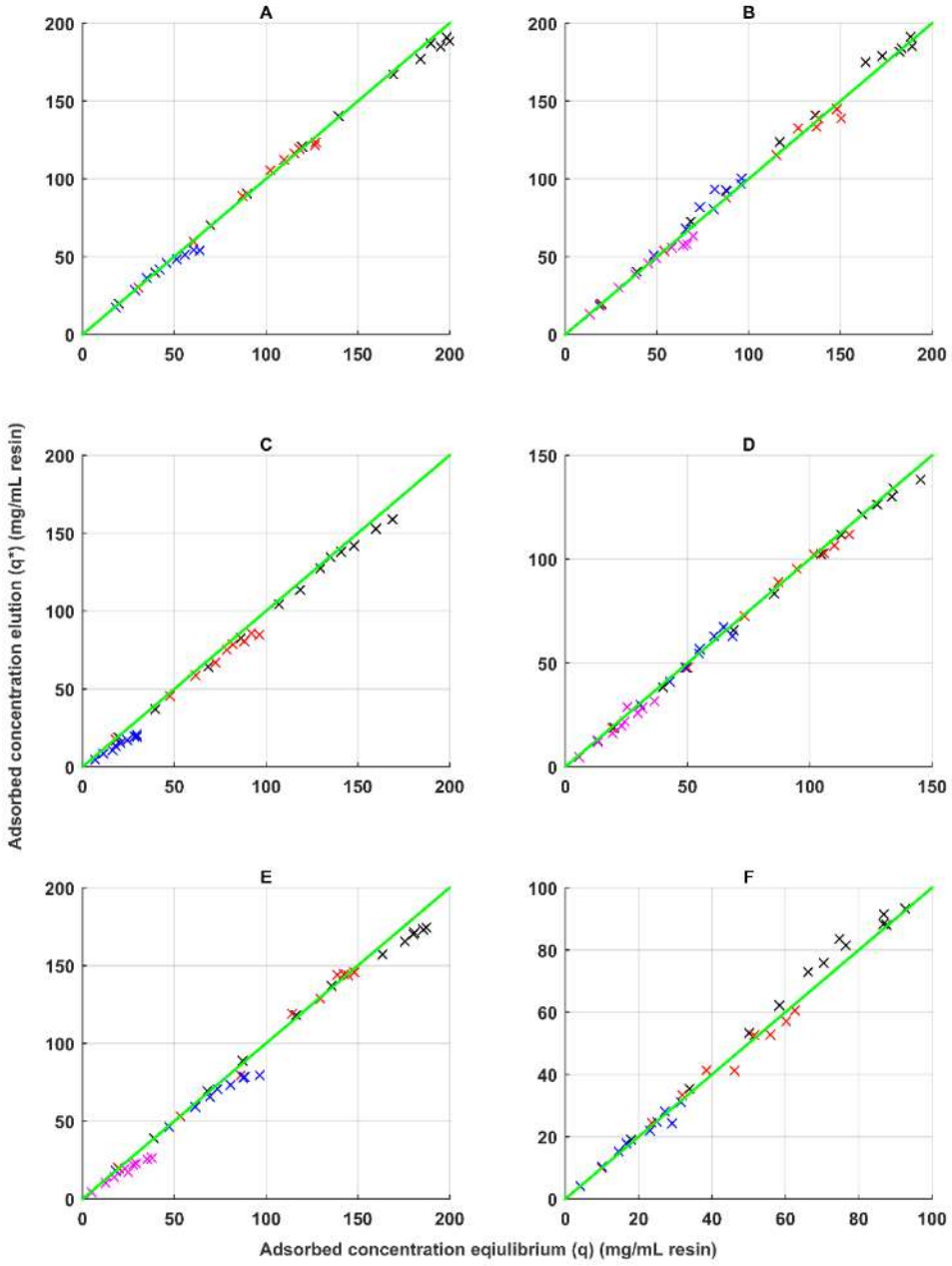

Figure 4-13 Comparison of adsorbed concentration calculated via equilibrium measurement and mass balance calculation versus via elution. Panels A, C & E show isotherm data generated at pH 9 and panels B, D & F at pH 8 50 mM Tris. Panels A & B show BSA isotherms, panels C & D show Ova isotherms and panels E & F show Con data. Different colours represent NaCl levels for isotherms displayed and are described in Table 4-5.

4.7 Discussion and summary

As outlined in the aims and objectives for this chapter critical factors which affect the data quality of SCIs have been tested and an experimental which delivers reliable results has been decided on. An effective sanitisation procedure which achieved negligible carry-over for the fixed tip ALHS was decided upon. Both protein and sanitisation fluid (0.5 M NaOH) carry-over was tested.

Comparability of isotherms across different phase ratios using PreDictor plates in strongly interacting conditions was demonstrated using different experimental variable. Comparability between PreDictor plates and beaker scale experiments was also demonstrated in strongly interacting conditions. Conversely, weakly interacting isotherms using 2 µL of chromatography resin were shown to compare poorly with 20 and 50 µL of resin. However, the use 20 and 50 µL of resin produced isotherms in good agreement when measured at equilibrium and via elution. Propagation of precision error in adsorbed concentration using different volumes of chromatography resin at equilibrium and elution suggested why 2 µL of chromatography resin may be a poor choice and that under certain conditions calculation of adsorbed concentration via elution can be preferable. In order to saturate increased resin volumes increased stock concentrations are required and the effect of those increased protein stock concentration on liquid handling accuracy was assessed and any inaccuracies resolved. Additionally, the choice of the three model proteins was discussed and their behaviour in column experiments was shown. SCIs of these model proteins was shown and their trends discussed before showing that error propagation of equilibrium concentration assayed directly was less than when equilibrium concentration was quantified via mass balance from elution. Finally a comparison between equilibrium data and elution data was done for all the SCIs discussed and some suggestion for ways to maximise agreement between the two methods were decided upon.

The expectation is that all the learning from this chapter can be applied to multicomponent isotherms, which are more challenging to study. It is therefore

hoped that this method will produce high-quality MCIs of proteins, which are reported in the literature as being difficult and time consuming to study.

Chapter 5 Quantification of protein mixtures

5.1 Aims and objectives of chapter

In order to generate multicomponent isotherm data, first a platform methodology to generate reliable data needs to established as discussed in Chapter 4. Secondly an analytical method which can reliably assay individual protein concentrations from a protein mixture is required, which is the subject of this chapter. Traditional HPLC methods for quantification were discussed in section 2.5 HPLC separation of model mixture, here an alternative solution for quantification is presented. The alternative method involves taking only UV spectra of the protein mixtures and building a statistical model using multivariate data analysis in order to estimate individual protein concentrations. The specific aims of this chapter are as follows:

- Present the use of UV spectra in conjunction with multivariate data analysis in order to quantify proteins.
- Presentation of a method to clean up data used to calibrate and validate the multivariate model and how to go about creating a robust model.
- Discussion of the performance of the multivariate models.

This chapter seeks to present a method which can be used to quantify protein mixtures in order to study multicomponent isotherms. Once the method has been established its performance will then be discussed.

5.2 Quantification of model mixture using UV spectra

Quantification of protein mixtures using UV spectra is an interesting alternative to HPLC. Whilst there is significant overlap in the spectra of BSA, Ova and Con which will require a deconvolution method in order to estimate individual concentrations from a mixture the analysis time required to generate a UV spectra for 96 samples is approximately 30 minutes making it a rapid analytical method well suited to multicomponent isotherm studies. As discussed in section 2.5 HPLC separation of model mixture any single HPLC method chosen will also require a deconvolution method, as such UV spectra in conjunction with a deconvolution method is an attractive alternative.

Different proteins will have different UV spectra dependant on the presence of aromatic residues such as tryptophan, tyrosine, and phenylalanine which have a local maxima around 280 nm. Additionally there is another local maxima around 230 nm due to the aliphatic chain as displayed in Figure 2-12 (Hansen et al., 2011). Figure 5-1 shows the raw UV absorbance spectra of the model proteins, Con shows the strongest absorbance across the spectra with BSA and Ova having similar absorbance levels to one another. The position of the maxima/minima and differences in the shape of the spectra can be seen in the normalised spectra. BSA and Ova have their maxima at the lowest wavelength at 243 nm with Ova nearly equalling the maxima again at 279 nm whereas Con's maxima is at 281 nm. There are subtle differences in the shape of the spectra throughout in the normalised spectra in Figure 5-1.

Figure 5-2 shows normalised spectra of BSA, Ova and Con at pH 8 and 9. Ova spectra seems stable across both pHs but there are shifts in both BSA and Con spectra across the pHs. It was therefore decided that pH 8 and 9 binary PLS models needed different calibration/validation samples and different PLS models in order to quantify samples at pH 8 and 9.

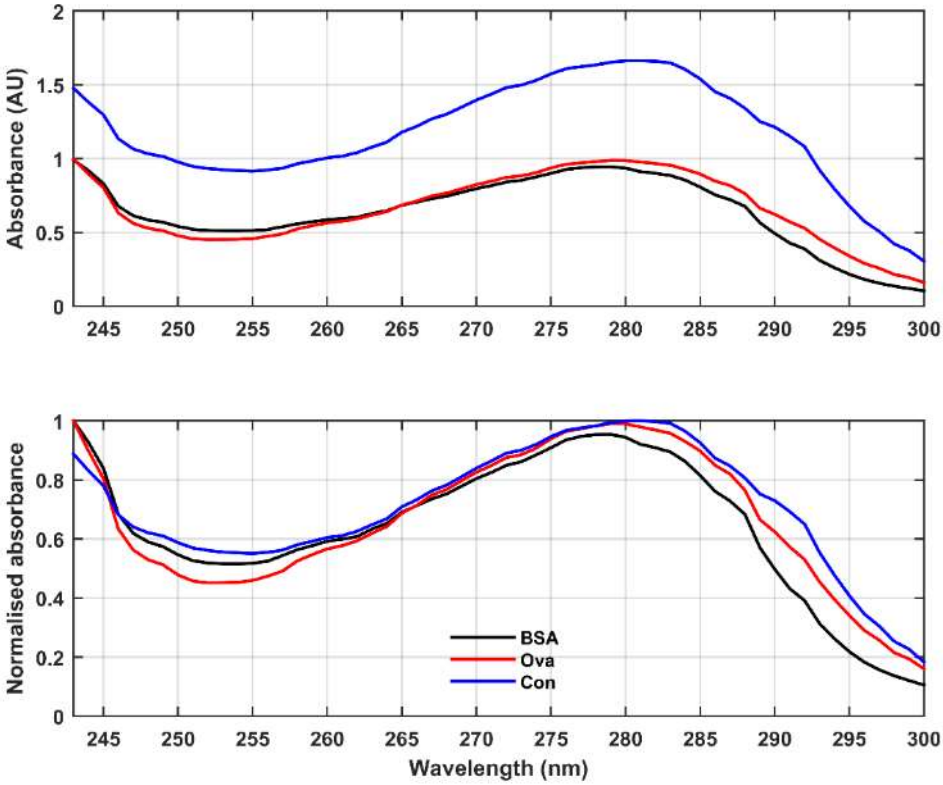

Figure 5-1 UV spectra of BSA, Ova and Con. Top panel shows raw absorbance of proteins at 3 mg/mL and bottom panel shows absorbance internally normalised to its maximum absorbance.

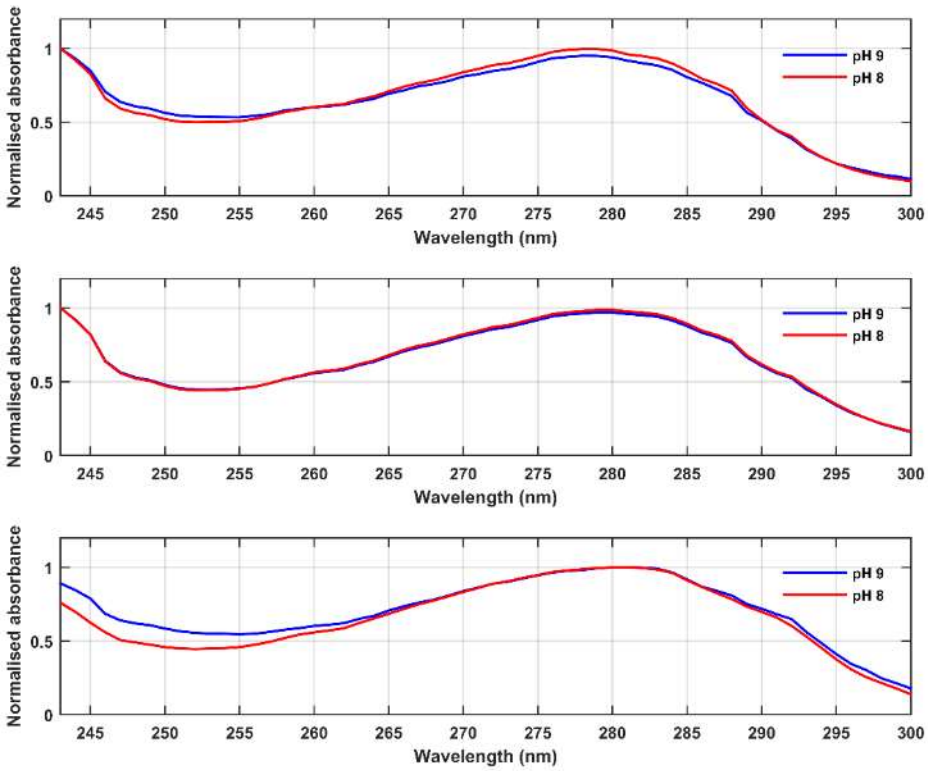

Figure 5-2 Normalised spectra of BSA, Ova and Con at pH 8 and and 9. Top plot is BSA, middle Ova and bottom Con.

5.2.1 Multivariate data analysis to quantify proteins from UV spectra

As shown in the section 5.2 Quantification of model mixture using UV spectra the spectra for BSA, Ova and Con are highly overlapping with subtle differences in maxima, minima and shape. A numerical or statistical method is required to leverage these differences in order to then estimate the individual protein concentrations in the mixture. The method chosen here is multivariate data analysis, specifically partial least squares regression also known as projection of latent structures (PLS).

PLS is a well suited to this problem as it can be applied to data sets with many, noisy, collinear variables in X and Y (Eriksson et al., 2013). During PLS model calibration samples of known composition are provided and the PLS model links variation in individual protein concentration to variation in the spectra. The spectra of samples with unknown composition are then provided and the PLS model predicts the individual concentration within the mixture.

The proteins in the model system being studied here are highly related albumins from different species which section 2.5 HPLC separation of model mixture showed are difficult to distinguish using analytical chromatography methods, other groups have also found these protein difficult to separate chromatographically (Kopaciewicz et al., 1983). The highly related nature of the albumins will likely increase the difficulty of quantification as well as potentially may make the system a more realistic replication of a chromatography polishing step where highly related product related impurities are at in being separated.

5.2.2 Preparation of binary control data sets of known concentration and spectra

In order to build the PLS models and test their efficacy a population of samples of know concentration is created. This is done in the same way that the feed plate is generated, stock solutions of the BSA, Ova, Con, diluent, NaCl and Tris stock at the required pH. Stock solutions of protein were prepared at concentrations of 20 mg/mL and 2 mg/mL in order to create mixtures of both high and low protein concentration whilst keeping the volume handled between 50 and 1000 µL where the liquid

handling capability of the ALHS is optimal as discussed in Chapter 4. Stock solutions were prepared gravimetrically but where quantified as single component solutions using standard curves so that all stock solutions were quantified to one standard curve and so were in agreement.

The choice for the range of proteins was chosen to firstly maximise the possible range for the mixtures of proteins. Previous examples of using PLS to quantity protein mixtures focused on rather narrow ranges of proteins typically between 0.1 and 1 mg/mL (Baumann et al., 2016; Brestrich, Briskot, Osberghaus, & Hubbuch, 2014; Dismer, Hansen, Oelmeier, & Hubbuch, 2013; Hansen et al., 2013, 2011). The work done here uses PLS models with much increased ranges from these. The rationale is that increased ranges improve the sensitivity of the assay. If the method can assay ranges for component A and B from 0.1 to 1 mg/mL but the sample in question is at a concentration of 2 mg/mL and 0.1 mg/mL for components A and B respectively then the sample will need to be diluted 2-fold in order to assay it in the calibrated space. After a 2-fold dilution the sample will be and a concentration of 0.05 mg/mL and 1 mg/mL for compounds A and B respectively meaning compound A is now below the lowest protein containing calibration sample, so by increasing the range of the assayable space for both compounds the sensitivity of the assay is improved. However, the range of the PLS models should not be so great that strongly absorbing wavelengths become saturated, meaning that increases in protein concentration no longer result in increases in absorbance. For this reason the BSA-Ova binary calibration was limited from 0-4 for BSA and 0-6 for Ova, a ceiling of 10 mg/mL total protein. In each of the single component BSA, Ova standard curves absorbance response flattens at approximately 10 mg/mL as shown in Figure 5-3. In the case of BSA-Con and Ova-Con binary PLS mixtures the range was chosen as 0-4 and 0-3 mg/mL respectively. These ranges were chosen because larger models including higher concentrations did not perform well (data not shown). A shorter path length using less volume in each well could extend the linear range of

absorbance. However, there is also a trade of in sensitivity to consider.

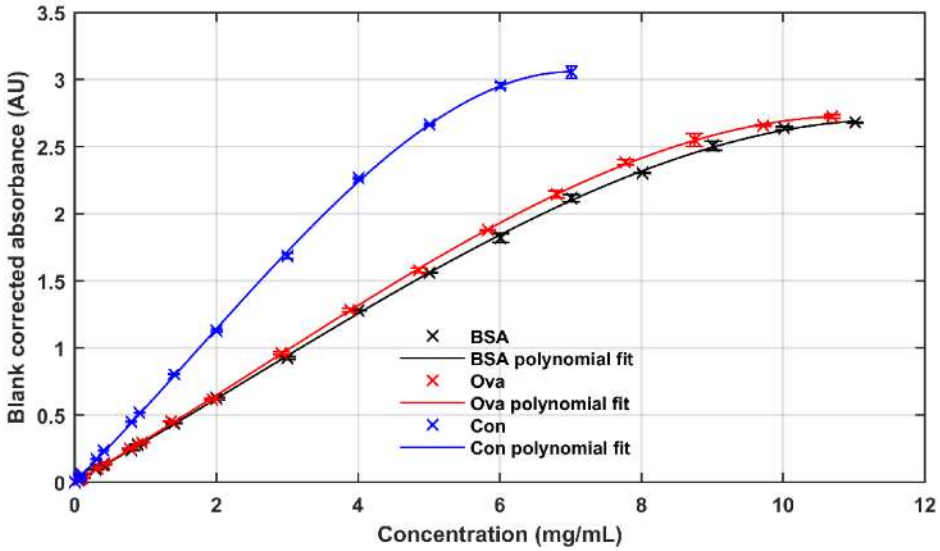

Figure 5-3 Standard curves for BSA, Ova and Con at 280 nm. Typically standard curves were used in the linear response range and linear descriptions were used to quantify protein, standard curves have been displayed here beyond this linear range for illustrative purposes.

The population of samples generated for calibration and validation of PLS models is shown in Figure 5-4. The protein mixtures shown in Figure 5-4 were prepared in triplicate across 4 96 well plates. Points where chosen in a space filling grid layout and are almost certainly over populated, the reason for this was firstly to have a robust data set so that models are not overly dependent on small number of points. Secondly as the ranges have been extended from what has been previously discussed in the literature it is useful to have a well populated space to allow assessment of assay performance throughout. Additionally the space below 1 mg/mL was populated more densely as it is likely that lower concentration of proteins will be more difficult to assay as their signal is smaller. Finally single component samples were included in the model as strongly interacting isotherms have a tendency increase in q whilst showing no increase in c as shown in Figure 4-11. In practical terms this may mean that sections of some multicomponent isotherms may contain significant amounts of single component data and the PLS model must be able to effectively assay in this single component space.

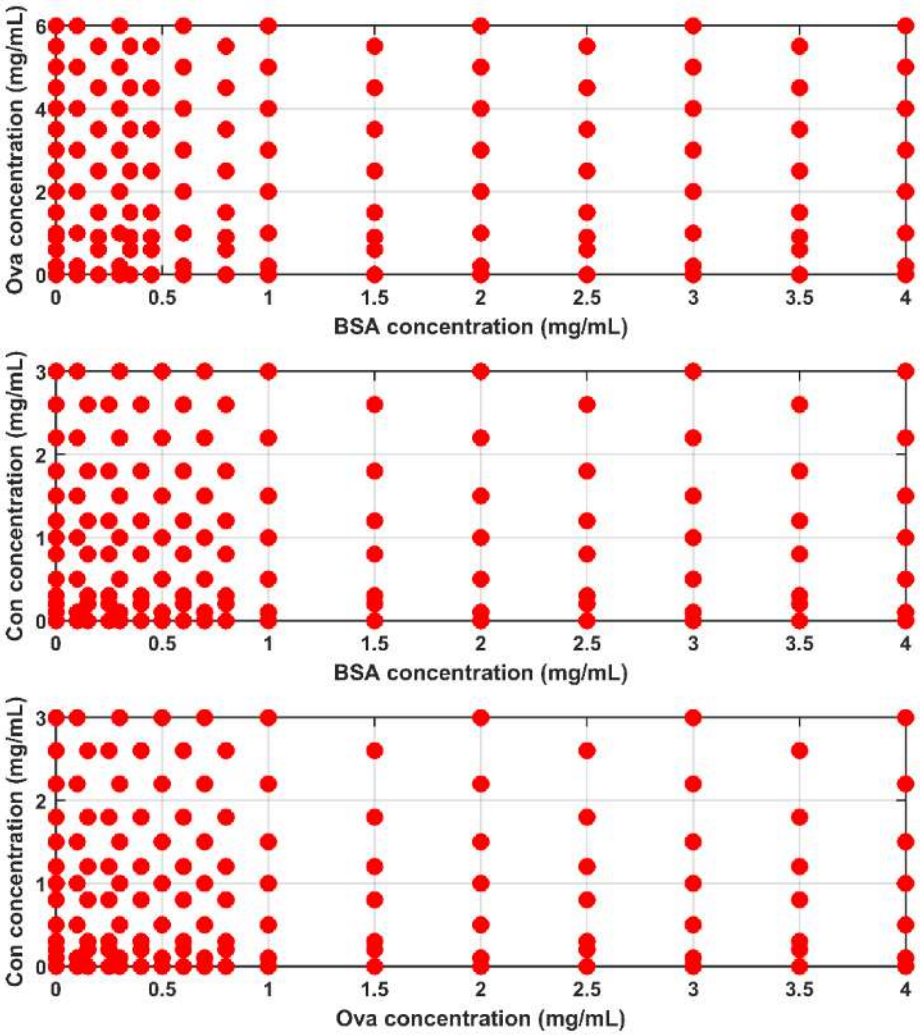

Figure 5-4 Concentration levels for calibration and validation of PLS models for all binary mixtures.

5.2.2.1 Outlier testing and rejection of spectra in PLS model building

In order to build PLS models of the best possible standard outlier detection and rejection was done at 2 levels. As previously mentioned PLS calibration and validation solutions were prepared in triplicate, these triplicate solutions were screened for an outlier before the remaining replicates were averaged. The method used for outlier rejection was median absolute deviation (MAD), described in equation 5-1, which relies on the median values and median deviation around the median instead of using average values and standard deviations. Both the mean and standard deviation can be strongly affected by the presence of just one outlier, whereas the median and median absolute deviation are more robust methods which is not affected by outliers

until 50% of the data become outlying (Leys et al., 2013; Rousseeuw & Hubert, 2011). This also means MAD is well suited to outlier detection of small data sets such as the triplicate spectra being explored here. The score is subsequently calculated and is described in equation 5-2, the limit for the score was set to 3 which is considered very conservative as opposed to 2 which is considered poorly conservative. A cut off of 2 would reject spectra which are closer to the median than a cut off of 3. The spectra being considered are from 243 nm to 300 nm at intervals of 1 nm meaning every wavelength in the spectra generates a score, spectra where ≥ 10% of the spectra was considered outlying were rejected and the remaining spectra replicates were averaged.

$$\mathrm{MAD} = 1.483 \cdot \mathrm{median}_i(|x_i - \mathrm{median}\ (x_j)|) \qquad 5\text{-}1$$

$$\mathrm{score} = \ (x_i - \mathrm{median}\ (x_j))/\mathrm{MAD} \qquad 5\text{-}2$$

Figure 5-5 show outlier rejection test in in practice, the examples is taken form BSA-Ova binary PLS calibration/validation samples pH 9. In the left hand plot the 3 spectra replicates can be seen with the outlier rejection boundary set at a score of 3 and in the right hand plot the averages before and after rejection. The left hand plot clearly shows a spectra which is substantially different from the other 2 and it is indeed identified as an outlier. The outlier boundary is clearly very tight for the spectra in the left hand plot. The high stringency means that many spectra will be rejected as is reflected in Table 5-1. However, this is not problematic as the remaining

replicates are always averaged meaning information is not lost by having tight boundaries and rejecting many replicate spectra.

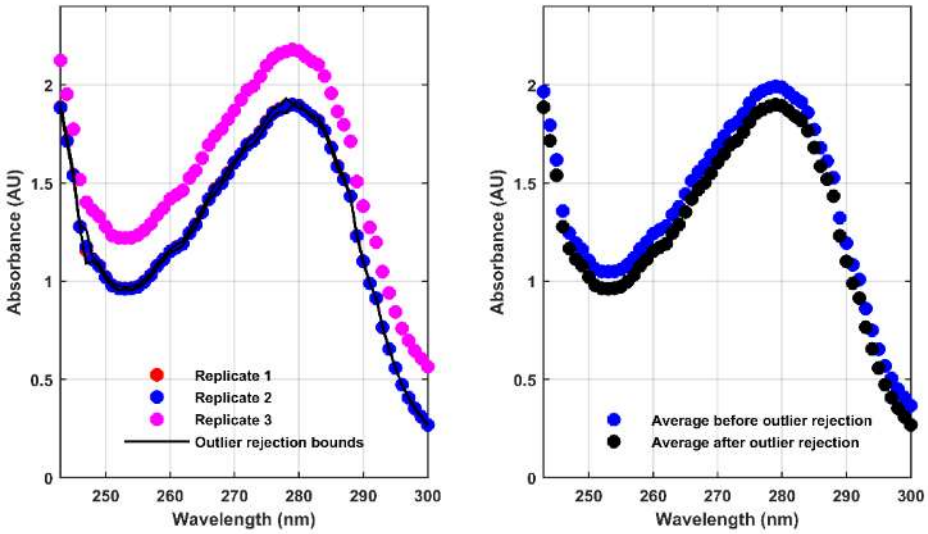

Figure 5-5 Spectra showing outlier rejection test on BSA-Ova mixture at pH 9. Left hand panel shows all 3 replicates and the outlier rejection boundary set at score of 3, right hand panel shows the averages before and after rejection.

Binary mixture	*pH 8*	*pH 9*
BSA-Ova	101	84
Ova-Con	111	94
BSA-Con	105	108

Table 5-1 Summary of the number of outlying spectra for each binary data set. The total number of samples in each data set was 384.

5.2.2.2 Outlier detection of binary multivariate datasets

Once the data set has been created, the spectra have had outliers rejected and averages taken they are supplied to SIMCA along with the concentrations of the individual components. An initial model is then built with the entire data set and at least 2 latent variables are added to the model. This model is used to detect outliers within the entire dataset. Outlying spectra have been detected and rejected explained in the previous section, but there still may be outlying samples when considering in the context of a multivariate model. These samples are detected manually using 3 plots: OLevY against YvarResST, Ypred against YvarRes and T2Range against DmodX, these abbreviations are explained in the below paragraph.

OLevY is a measure of the influence of a point (observation) on the PLS model in the Y space and YvarResST is the residual in the concentration estimate displayed in standardised units where the residual is divided by the standard deviation. This plot is giving information about how well an individual sample is predicted by the model and also how important that sample is in determining the model. Samples with high YvarResST should be considered for rejection and samples with high YvarResST and a high OLevY should certainly be rejected as they are poorly estimated by the model and have a large influence on the model, an example of the plot is shown in Figure 5-6. Samples with low YvarResST and high OLevY may also be troublesome as they are affecting the model rather a lot. Hottelling's T2Range is a distance measure for how far away point lie from the centre of a model on the hyperplane. DModX is the distance of a point from the hyperplane of the model. Severe outliers can be detected with high T2Range values, and moderate outlier can be detected with high DmodX values. Points which lie away from the population in either of these metrics should be considered for rejection as shown in Figure 5-7. Ypred is the predicted concentration and YVarRes is the residual in prediction. YVarRes values which lie away from the bulk of local data should be rejected as they are not predicted well by the model and so may be errant as shown in Figure 5-8.

Figure 5-6, Figure 5-7 and Figure 5-8 show that point 117 lies away from all the other data points for all 3 outlier test. As such it was omitted from models before

any further model building was done in the BSA-Ova pH 8 PLS model. 3 other points also lie away from the bulk of data in Figure 5-7 but there were left in the model as they are not outlying using the other metrics. Table 5-2 shows the number of outliers detected using these methods in each binary PLS model.

Binary mixture	*pH 8*	*pH 9*
BSA-Ova	1	2
Ova-Con	1	0
BSA-Con	2	0

Table 5-2 Number of outliers detected using multivariate analysis for each binary PLS model. Total number of samples is 128.

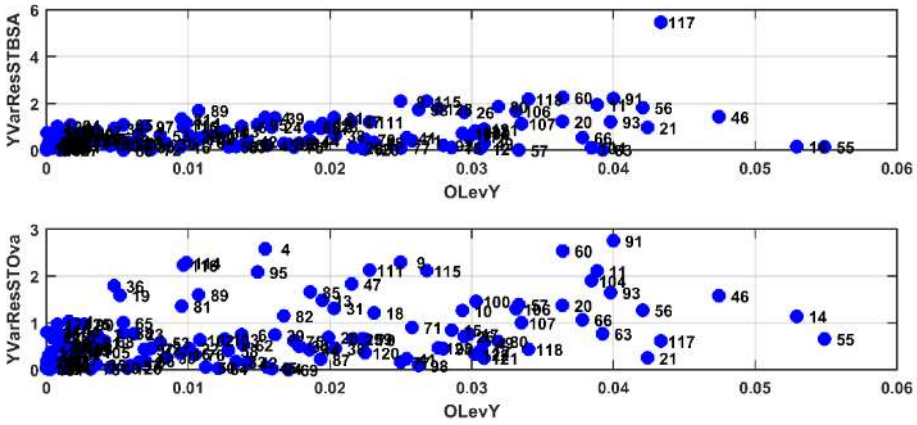

Figure 5-6 Plotting OLevY (influence of point on model) against YVarResST (scaled residual concentration estimate) for each component in the model to detect outliers. BSA-Ova model pH 8.

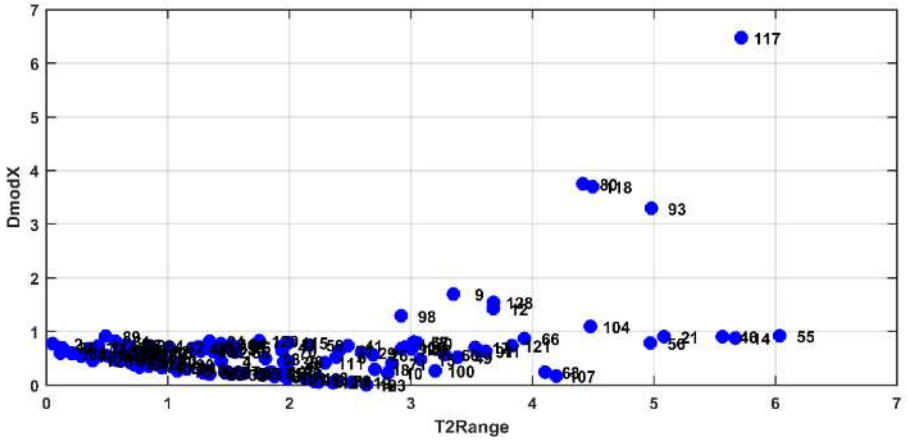

Figure 5-7 Plotting T2Range (distance from model centre on hyperplane) against DModX (distance from model hyperplane). BSA-Ova model pH 8.

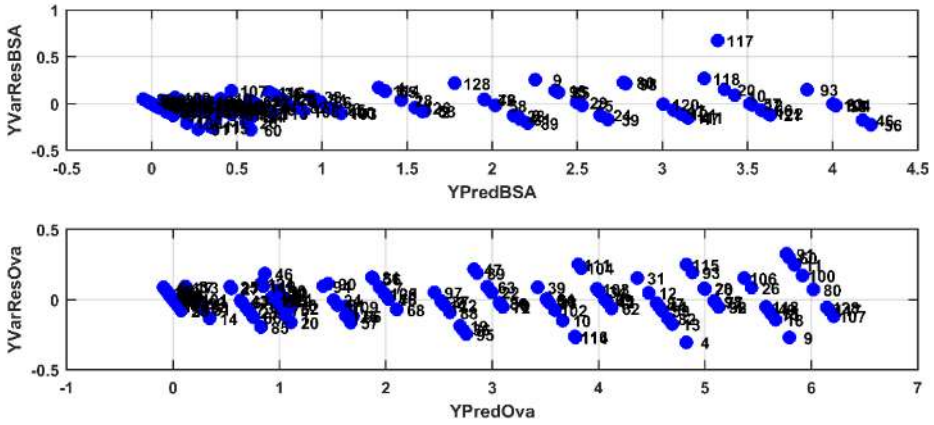

Figure 5-8 YPred which is predicted concentration against residual in prediction for each component in the model. BSA-Oba pH 8.

5.2.3 Binary PLS model building

5.2.3.1 Calibration validation split

Once the dataset has had spectral and multivariate outliers rejected, it is split into calibration and validation sub-sets. This is done by sorting the initial PLS model with all the data points by their t-score for the first LV and placing alternating samples in calibration and validation data sets. The reasoning behind this is that almost all the information is contained in the 1st LV and by splitting alternating samples between calibration and validation the information in the total data set is split somewhat evenly. An example of the resultant calibration validation split displaying concentration combinations is displayed in Figure 5-9.

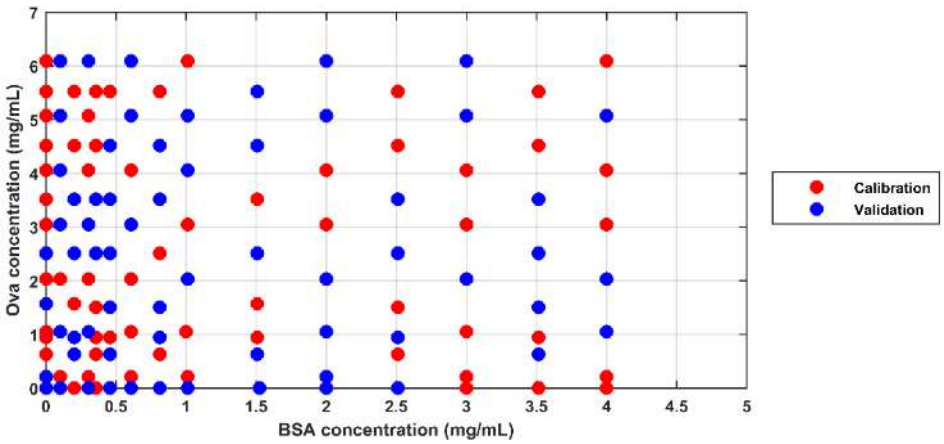

Figure 5-9 Sample split between sampled for calibration and validation. BSA-Ova data set at pH 8.

5.2.3.2 Choosing the number of latent variables

Choosing the number of LVs is done by monitoring the root mean squared error during cross validation (RMSEcv) as the number of LVs is increased. Cross validation is a standard practice when building multivariate models and as such it is well established (Eriksson et al., 2013; Wold, 1978). Cross validation is performed only on the calibration data set, the validation set only being used to assess the model once the model has been built. During cross validation many models are created from the calibration dataset by leaving a portion of the calibration data out (in SIMCA 7 models are created leaving 1/7th of the dataset out every time). The RMSEcv is then

calculated for all the models by predicting the left out portion of the data set for each model. As more LVs are added to a PLS model both the signal and the noise are being included in the model, the danger of adding too many LVs is that the models begin to fit too much of the noise rather than the signal. During cross-validation models are built with reduced data sets; if too much noise has been fitted in the model when the excluded data is predicted the model will perform worse when compared to a smaller number of LVs. In this way it is possible to choose a suitable number of LVs, there is either an increase in the RMSEcv as the LVs increase or there is negligible improvement. It generally best to use the smallest number of LVs possible (Wold, 1978).

An example showing how the number of LVs is chosen for BSA-Ova pH 8 PLS model in Figure 5-10. Table 5-3 compares how RMSEcv changes as more LVs are added. For this model 2 LVs were chosen, although looking at Figure 5-10 and Table 5-3 there is a minima in RMSEcv at 5 LV there is little additional benefit in adding more than 2 LVs, and as previously mentioned a simpler model is preferred. The number of LVs used to build PLS models for at binary PLS models is displayed in Table 5-4. All the binary models have between 2 and 3 LVs, this is in good agreement with similar work in other groups (Brestrich et al., 2014).

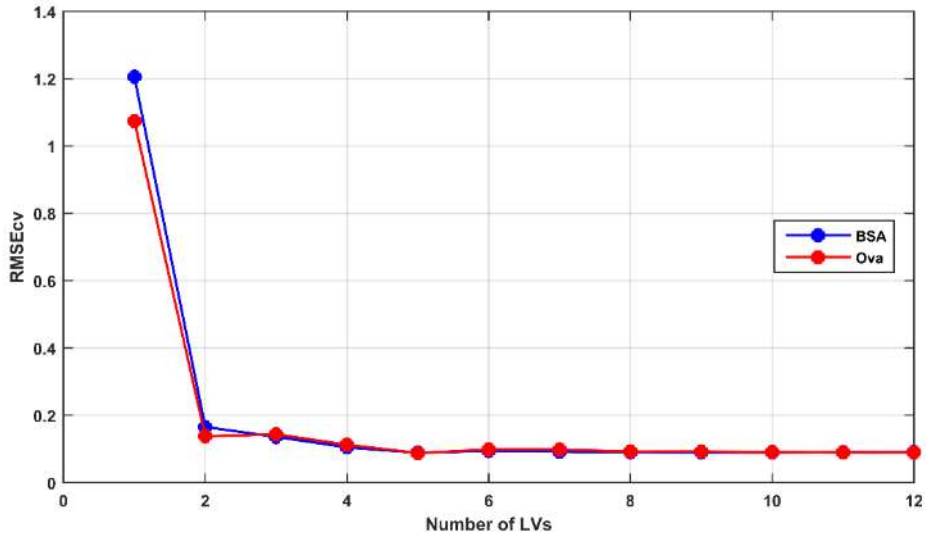

Figure 5-10 RMSEcv for BSA and Ova PLS model building for BSA-Ova pH 8.

	RMSEcv (mg/mL)		*Improvement from 1 LV*		*Improvement previous number of LV*	
LV number	**BSA**	**Ova**	**BSA**	**Ova**	**BSA**	**Ova**
1	1.20546	1.07365	0%	0%	N/A	N/A
2	0.165175	0.137725	-86%	-87%	-86%	-87%
3	0.135838	0.143471	-89%	-87%	-18%	4%
4	0.104617	0.112463	-91%	-90%	-23%	-22%
5	0.088614	0.087739	-93%	-92%	-15%	-22%
6	0.094267	0.098289	-92%	-91%	6%	12%
7	0.092151	0.098081	-92%	-91%	-2%	0%
8	0.090173	0.092255	-93%	-91%	-2%	-6%
9	0.089645	0.092487	-93%	-91%	-1%	0%
10	0.089372	0.090871	-93%	-92%	0%	-2%
11	0.090418	0.089557	-92%	-92%	1%	-1%
12	0.090996	0.0897	-92%	-92%	1%	0%

Table 5-3 Showing RMSEcv and percent change in RMSEcv for BSA-Ova PLS model at pH 8. RMSEcv has been highlighted, green colours represent lower RMSEcv and red higher values.

Binary mixture	*pH 8*	*pH 9*
BSA-Ova	2	2
Ova-Con	2	3
BSA-Con	3	3

Table 5-4 Number of latent variables used for each binary PLS model.

5.2.4 Binary PLS performance at predicting concentration

The performance of the PLS method will be discussed for each of the binary models. Only the performance of the pH 9 models will be discussed for brevity as pH 8 models behaved in a similar way. These assessments were done by requesting the model to generate concentration estimates for samples of known concentration.

5.2.4.1 BSA-Ova binary PLS model performance

The performance of the calibration and validation samples of the PLS model are displayed in Figure 5-11. Figure 5-11A and B are parity plots for BSA and Ova estimation respectively; they assess the linearity of the estimates as well as checking for bias which is assessed by fitting a linear regression to the data with the equation displayed in the plot. Figure 5-11A and B suggest that the model predictions remain linear across the range of the model despite the increased range used compared to other work (Hansen et al., 2013). There additionally looks to be very little bias in the model as the slopes are close to 1 and the intercepts are near 0.

Plots Figure 5-11C and D show the error in prediction for BSA and Ova respectively. The errors look in reasonable control mostly remaining within 0.1 mg/mL for BSA and 0.2 mg/mL for Ova. The data sets are fairly homoscedastic, there is a slight tendency for the errors to increase with increasing concentration, but when considered in terms of relative error (equation 5-3, all concentrations in mg/mL) this is not problematic as the increased denominator will mean relative error is negligible.

Figure 5-11E and F show display the dependency of relative error upon nominal BSA and Ova concentrations in order to assess the dependency of BSA error (or Ova error) on BSA and Ova concentration, BSA error is assessed in plot E and Ova plot F. Figure 5-11E suggests that there is some dependency of BSA error on Ova concentration on, when BSA concentration is low (<1 mg/mL) relative error is increased, this is unsurprising as the denominator becomes controlling in equation 5-3. In this region BSA relative error also increases as the Ova concentration increases. A possible explanation is that the sample has low BSA, high Ova signal meaning BSA contributes only a small input to the total signal making quantification challenging. The equivalent plot for Ova error in Figure 5-11 shows similar behaviour in that relative error increases as Ova concentration decreases with the denominator in equation 5-3 becoming controlling. However, the Ova relative error shows no dependence on the BSA concentration, possibly due to the ranges picked for the PLS model. Figure 5-3 shows that BSA and Ova present similar absorbance dependence on concentration in mg/mL with similar slopes to their standard curve. However, the PLS model was built with a greater range for Ova meaning BSA is quantified in more challenging situations. As previously mentioned the rational for this is that Ova is the less retained species on Capto^{TM} Q meaning that at equilibrium the assayed concentration of Ova on average will be greater than BSA. It was decided that some dependence of BSA error on Ova concentration is a reasonable sacrifice for increased assay range.

The errors associated with calibration and validation data sets seems quite comparable across both proteins. This is promising as it suggests the models have not been over fit: If too many LVs had been added to the model it might be expected that the validation dataset would perform worse than the calibration dataset.

$$\text{Relative error} = \frac{(\text{Predicted concentration} - \text{Nominal concentration})}{\text{Nominal concentration}} \quad \textit{5-3}$$

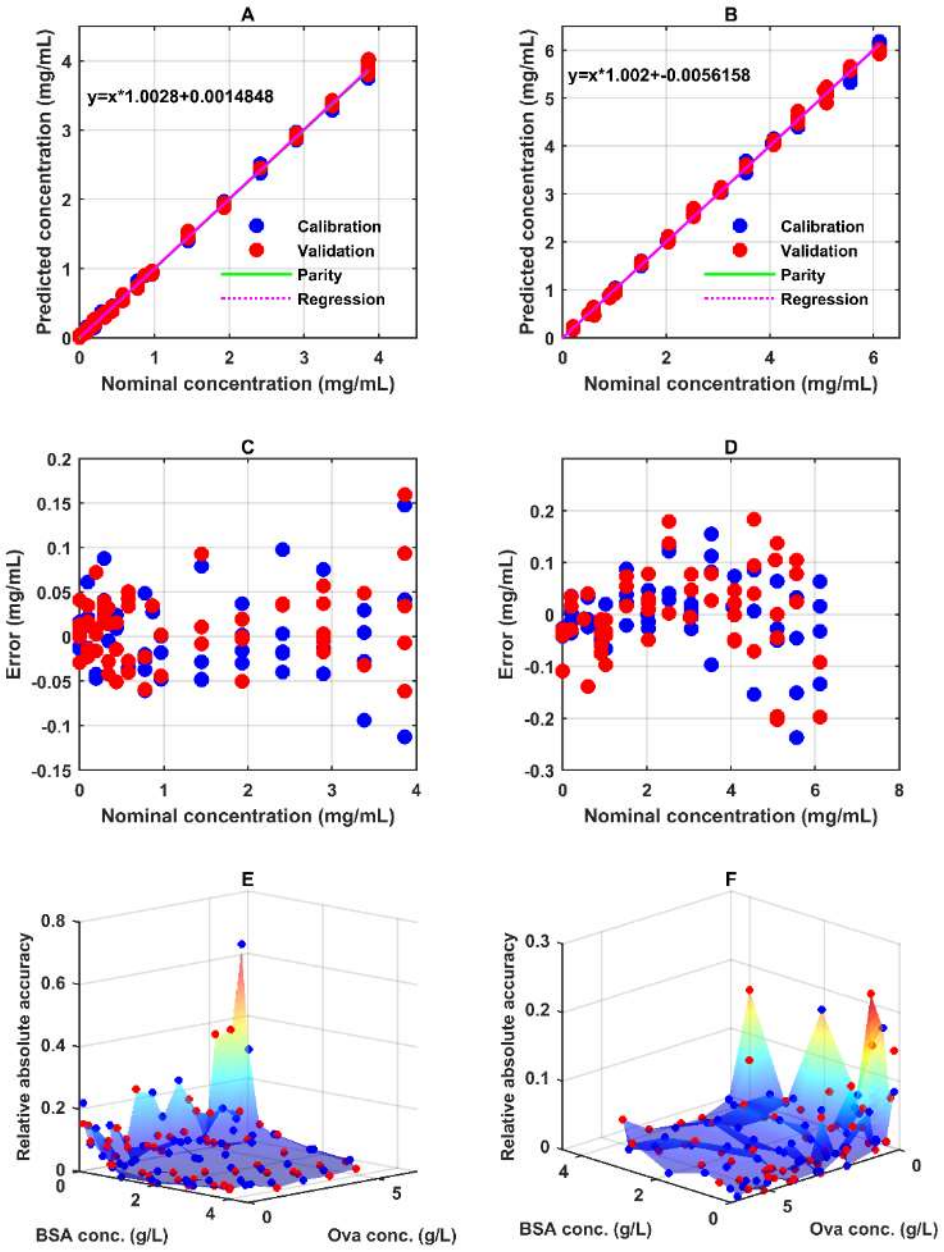

Figure 5-11 Performance of PLS models predicting concentration of individual components in binary mixtures of BSA-Ova at pH 9. Panels A, C and E refer to the prediction of BSA and panels B, D and F refer to prediction of Ova in the binary mixture. A and B are parity plots of nominal vs predicted concentration, a linear regression has been fit to the data to assess bias/linearity. Plots C and D assess the error in prediction assessing general performance. Plots E and F assess absolute relative error against BSA and Ova (equation 5-3) nominal concentration to assess the error dependence of one concertation on another.

5.2.4.2 Ova-Con binary PLS model performance

The performance of the calibration and validation samples of the PLS model are displayed in Figure 5-12. Figure 5-12A and B are parity plots for Ova and Con estimation respectively. Figure 5-12A and B suggest that the model predictions remain linear in this model as well. There additionally looks to be very little bias in the model as the slopes are almost equal to 1 and the intercepts are nearly 0.

Plots Figure 5-12C and D show the error in prediction for Ova and Con respectively. The errors look in reasonable control mostly remaining within 0.2 mg/mL for Ova and 0.1 mg/mL for Con. The data sets are fairly homoscedastic; there is a slight tendency for the errors in to increase with increasing concentration, when considered in terms of relative error (equation 5-3, all concentrations in mg/mL) this is not problematic as the increased denominator will mean relative error is negligible.

Figure 5-12E and F show display the dependency of relative error upon nominal Ova and Con concentrations in order to assess the dependency of Ova error (or Con error) on Ova and Con concentration, Ova error is assessed in plot E and Con plot F. Figure 5-12E suggests that there is some dependency of Ova error on Con concentration on, when Ova concentration is low (<1 mg/mL) relative error is increased, this is unsurprising as the denominator becomes controlling in equation 5-3. In this region Ova relative error also increases as the Con concentration increases. This is likely because in these circumstances of low Ova, high Con the signal presented by Ova is small relative to the total signal making quantification challenging. The equivalent plot for Con error in Figure 5-12 shows the equivalent behaviour with relative error increasing as the nominal concentration decreases and the denominator becomes controlling. There is also a dependency of Con error on Ova concentration, it increases as the Ova concentration increases in the low Con region. These range for the PLS models was chosen for the following reasons: Ideally the range for Con would be greater than Ova because Con is the less retained solute on Capto^{TM} Q. However, Figure 5-3 shows that Con has greater absorbance per mg/mL than Ova. Models with a greater range for Con behaved poorly (data not displayed) and the model presented was the largest that performed acceptably. It

was decided that some dependence of Ova error on Con, and Con error on Ova concentration is a reasonable sacrifice for increased assay range.

The prediction of Ova concentration was improved with the use of local models: Models were built which were decreased in range for Ova, running from 0-1 mg/mL for Ova and 0-3 mg/mL for Con (the range for Con was unchanged), this model was then used to predict the concentration of Ova in the range 0-1 mg/mL for Ova. When the predicted concentrations of Ova fell within the range 0-1 mg/mL using the local model the local model is used to predict the Ova concentration. When the local model predicted the concentration > 1 mg/mL for Ova the global model was used to predict the Ova concentration, the global model was always used to predict Con concentration. The performance this method is assessed in Figure 5-13 and it shows decreased error in the range 0-1 mg/mL of Ova when comparing plots Figure 5-12C with Figure 5-13B.

The errors associated with calibration and validation data sets seems quite comparable across both proteins. This is promising as it suggests the models have not been over fit, if too many LVs had been added to the model it might be expected that the validation dataset would perform worse than the calibration dataset.

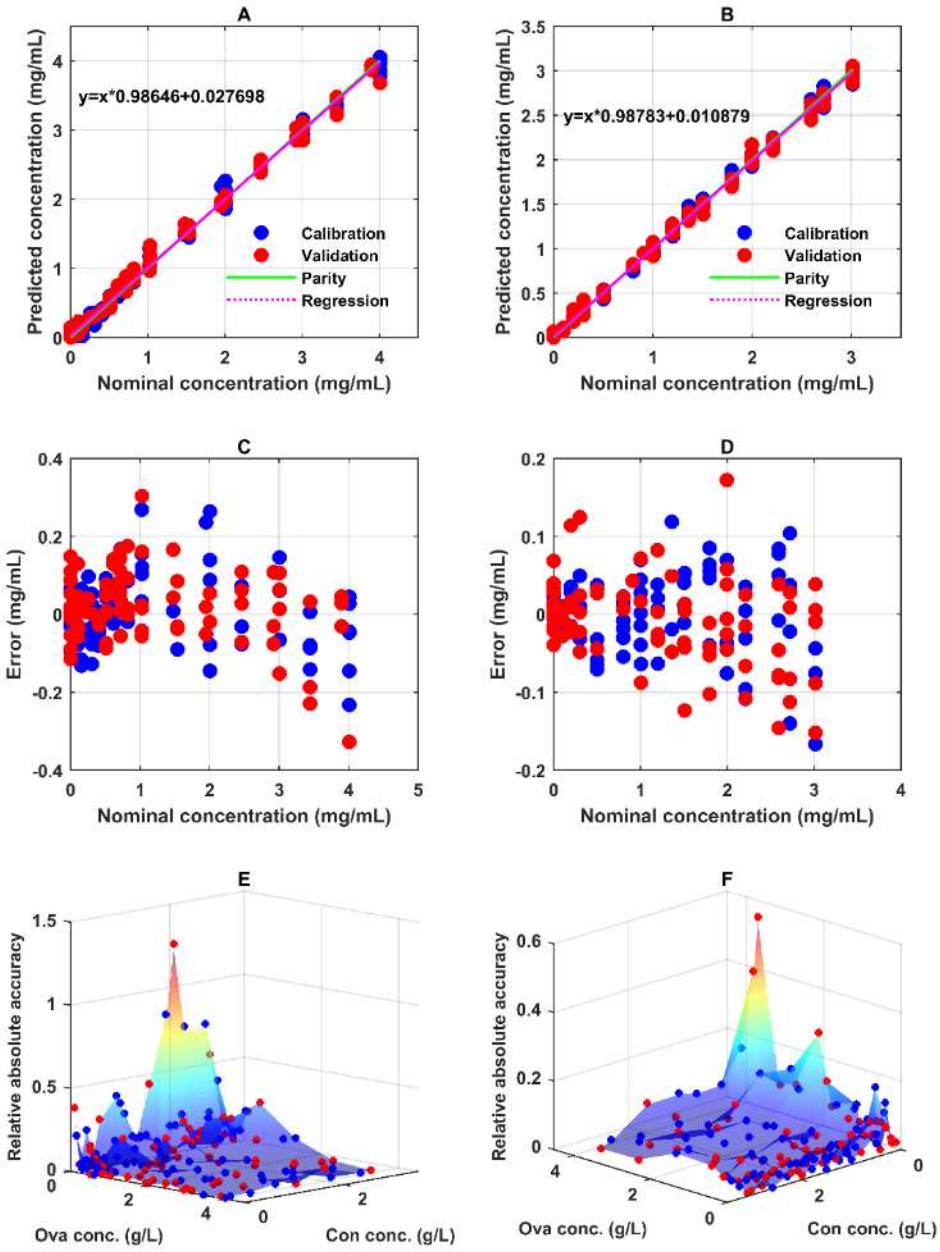

Figure 5-12 Performance of PLS models predicting concentration of individual components in binary mixtures of Ova-Con at pH 9. Panels A, C and E refer to the prediction of Ova and panels B, D and F refer to prediction of Con in the binary mixture. A and B are parity plots of nominal vs predicted concentration, a linear regression has been fit to the data to assess bias/linearity. Plots C and D assess the error in prediction assessing general performance. Plots E and F assess absolute relative error against Ova and Con (equation 5-3) nominal concentration to assess the error dependence of one concertation on another.

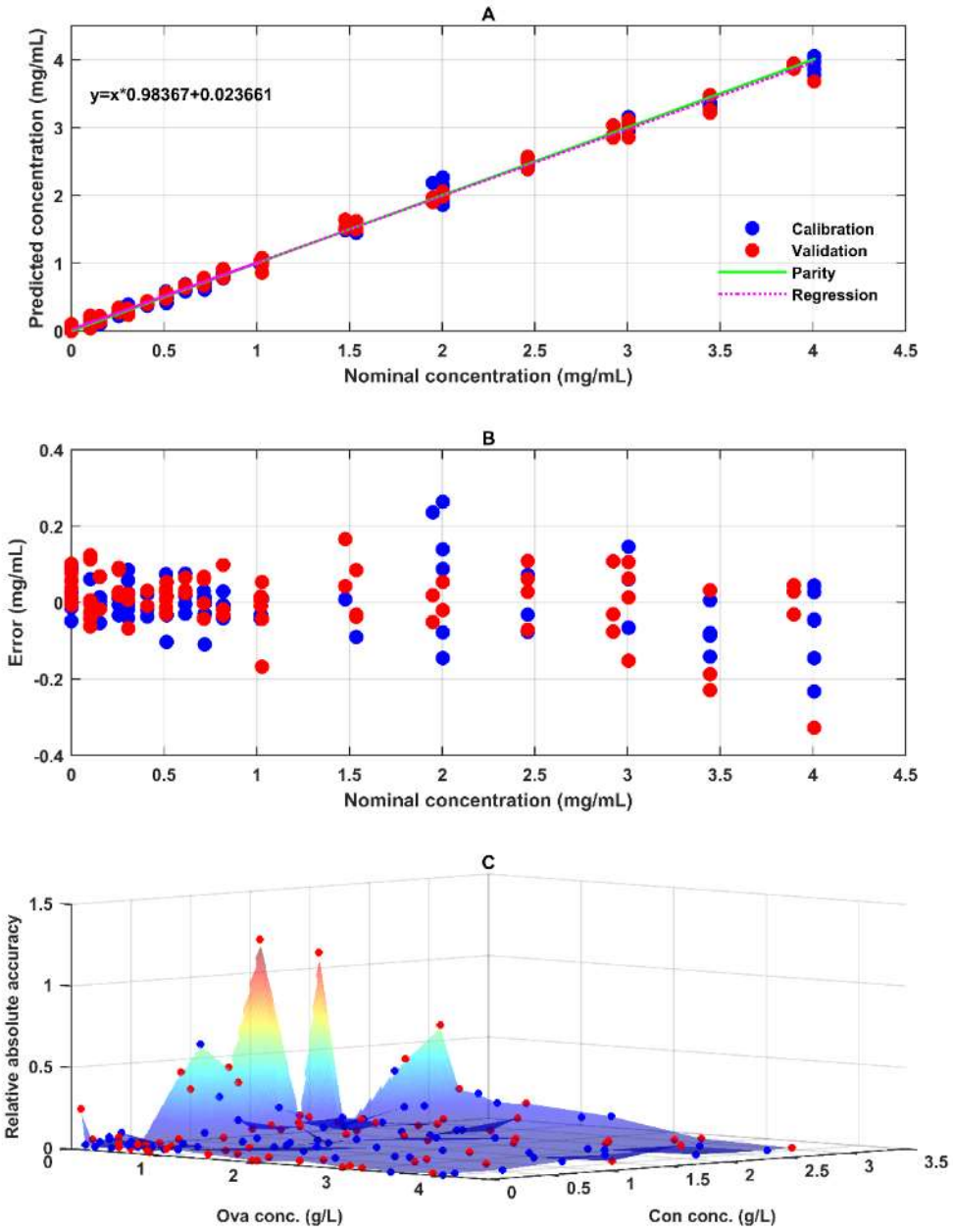

Figure 5-13 Performance of PLS models predicting Ova concentration in binary mixtures of Ova-Con at pH 9, predictions between 0-1 mg/mL of Ova have been done using a local model, other prediction used the global model. A is a parity plot of nominal vs predicted concentration, a linear regression has been fit to the data to assess bias/linearity. Plot B assesses the error in prediction assessing general performance. Plot C assesses absolute relative error (equation 5-3) against Ova and Con nominal concentration to assess the error dependence of Ova prediction on Ova and Con concentrations.

5.2.4.3 BSA-Con binary PLS model performance

The performance of the calibration and validation samples of the PLS model are displayed in Figure 5-14. Figure 5-14A and B are parity plots for BSA and Con estimation respectively. Figure 5-14A and B suggest that the model predictions remain linear in this model as well. There additionally looks to be minimal bias in the model as the slopes are almost equal to 1 and the intercepts are near 0.

Plots Figure 5-14C and D show the error in prediction for BSA and Con respectively. The errors look in reasonable control mostly remaining within 0.2 mg/mL for BSA and 0.1 mg/mL for Con. The data sets are fairly homoscedastic; there is a slight tendency for the errors in Con to increase with increasing concentration. But this is not problematic when considered in terms of relative error (equation 5-3, all concentrations in mg/mL) as the increased denominator will cause the relative error to be negligible.

Figure 5-14E and F show display the dependency of relative error upon nominal BSA and Con concentrations in order to assess the dependency of BSA error (or Con error) on BSA and Con concentration, BSA error is assessed in plot E and Con plot F. Figure 5-11E suggests that there is some dependency of BSA error on Con concentration on, when BSA concentration is low (<1 mg/mL) relative error is increased, this is unsurprising as the denominator becomes controlling in equation 5-3. In this region BSA relative error also increases as the Con concentration increases. This is likely because in these circumstances of low BSA, high Con the signal presented by BSA is small relative to the total signal making quantification challenging. The equivalent plot for Con error in Figure 5-11 shows the equivalent behaviour with relative error increasing as the nominal concentration decreases and the denominator becomes controlling. There is also a dependency of Con error on BSA concentration, it increases as the BSA concentration increases in the low Con region. These range for the PLS models was chosen for the same reasons discussed in *5.2.4.2 Ova-Con binary PLS model performance*: Ideally the range for Con would be greater than BSA because Con is the less retained compound on Capto^{TM} Q. However, Figure 5-3 shows that Con has greater absorbance per mg/mL than BSA. Models with

a greater range for Con failed (data not displayed) and this model was the largest did work. It was decided that some dependence of BSA error on Con, and Con error on BSA concentration is a reasonable trade-off for increased assay range. Similar to the previously discussed binary PLS model the prediction of Ova concentration was improved with the use of local models: Models were built which were decreased in range for BSA, running from 0-1 mg/mL for BSA and 0-3 mg/mL for Con (the range for Con remained unchanged), this model was then used to predict the concentration of BSA in the range 0-1 mg/mL for BSA. When the predicted concentrations of BSA fell within the range 0-1 mg/mL using the local model the local model is used to predict the BSA concentration. When the local model predicted the concentration > 1 mg/mL for BSA the global model was used to predict the BSA concentration, the global model was always used to predict Con concentration. The performance this method is assessed in Figure 5-15 and it shows decreased error in the range 0-1 mg/mL of Ova when comparing plots Figure 5-14C with Figure 5-15B.

The errors associated with calibration and validation data sets seems quite comparable across both protein. This is promising as it suggests the models have not been over fit; if too many LVs had been added to the model it might be expected that the validation dataset would perform worse than the calibration dataset.

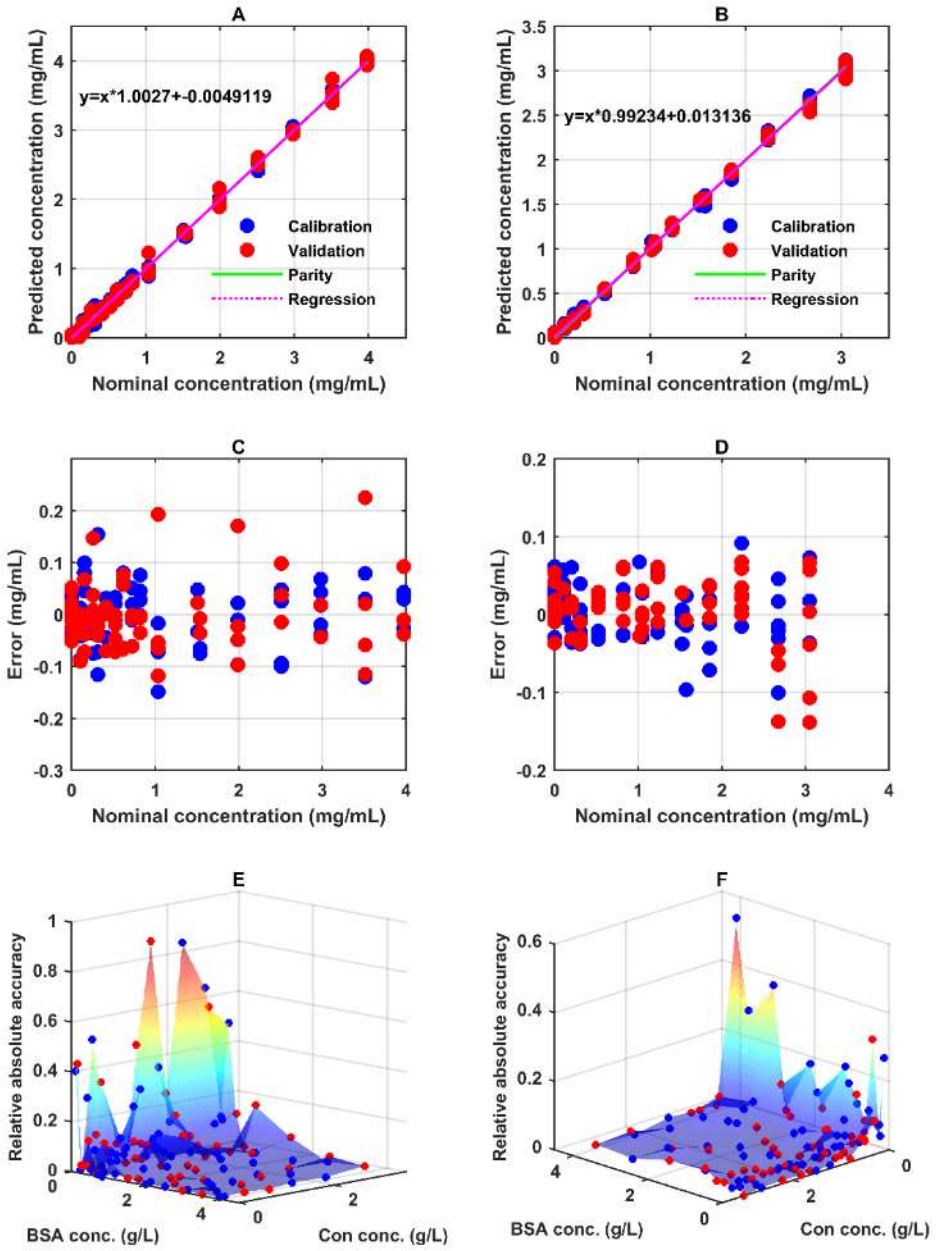

Figure 5-14 Performance of PLS models predicting concentration of individual components in binary mixtures of BSA-Con at pH 9. Panels A, C and E refer to the prediction of BSA and panels B, D and F refer to prediction of Con in the binary mixture. A and B are parity plots of nominal vs predicted concentration, a linear regression has been fit to the data to assess bias/linearity. Plots C and D assess the error in prediction assessing general performance. Plots E and F assess absolute relative error against BSA and Con (equation 5-3) nominal concentration to assess the error dependence of one concertation on another.

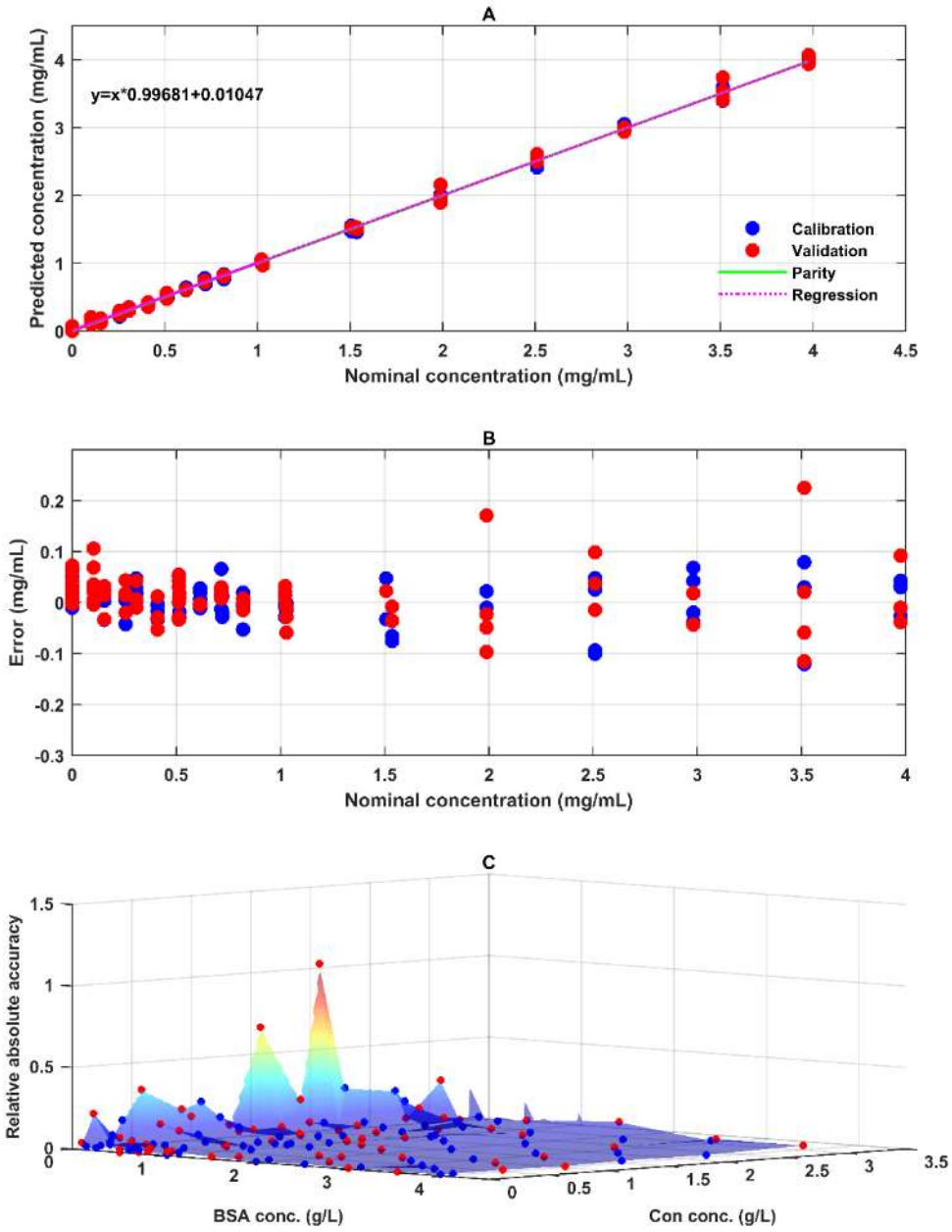

Figure 5-15 Performance of PLS models predicting BSA concentration in binary mixtures of BSA-Con at pH 9, predictions between 0-1 mg/mL of BSA have been done using a local model, other prediction used the global model. A is a parity plot of nominal vs predicted concentration, a linear regression has been fit to the data to assess bias/linearity. Plot B assesses the error in prediction assessing general performance. Plot C assesses absolute relative error (equation 5-3) against BSA and Con nominal concentration to assess the error dependence of BSA prediction on BSA and Con concentrations.

5.2.5 Preparation of ternary control data sets of known concentration and spectra

Ternary PLS models were only built at pH 9, pH 8 models were not built for reasons discussed in Chapter 6. Just as for the binary PLS models, sample of known concentration were generated and their spectra collected in order to calibrate and validate the ternary models. The ternary dataset will include single component, as well as the binary solutions used previously to build binary PLS models in addition to ternary protein solutions. The ternary mixtures of BSA, Ova and Con were prepared in twenty 96-deep-square-well plates. Both single component and binary mixtures were included because it is likely the ternary isotherms will include singe component, binary and ternary solutions at equilibrium Figure 4-11. Shows that adsorbed concentration can rise rapidly with no increase in liquid phase concentration in conditions of strong interaction. This behaviour is also likely to reoccur in both binary and ternary isotherms meaning the PLS models required in order to assay solutions must be able to cope with such situations.

The concentrations of the samples generated as control solutions for PLS calibration and validation are shown in Figure 5-16. Samples were chosen in in a space filling layout. The ranges of protein concentration were 0-2 mg/mL for BSA and Ova and 0-3 for Con. The reasoning for this range is that models of the range 0-4 BSA/Ova 0-3 Con worked well in the binary case. As BSA and Ova have similar absorbance in terms of absorbance (AU)/concentration (mg/mL) as suggested by Figure 5-3 keeping a total range of 0-4 of BSA + Ova concertation seemed reasonable in order to achieve models of comparable performance. The 0-1 range for all 3 proteins is extremely well populated, this was done in case the chosen range was found to be too great and performance was poor, if this was the case then the ternary PLS model could easily be scaled back to 1 mg/mL for each component which other work has suggested would work well (Hansen et al., 2011). One consequence of this is that the space from 1-2 mg/mL for BSA and Ova is less well populated. It will be shown later that this space is well quantified despite the comparatively disparate population of samples.

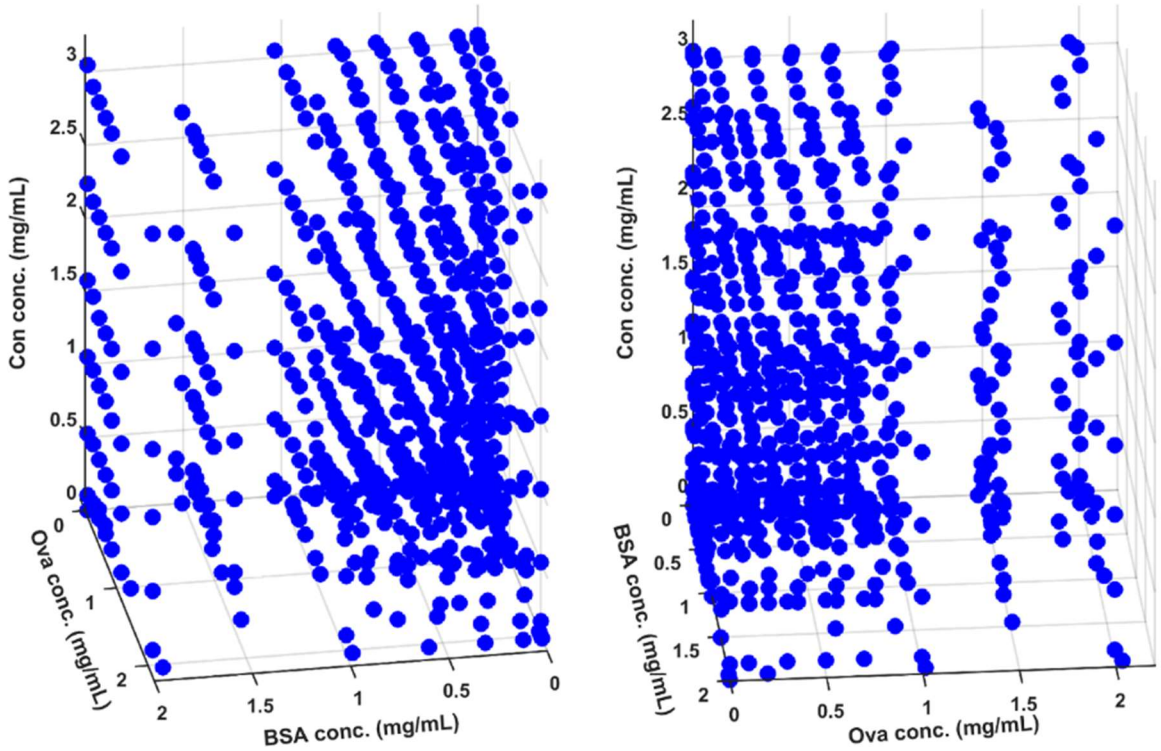

Figure 5-16 3D plot of control samples involved in calibration and validation of 3 component PLS models for the quantification of individual concentrations from a 3 component mixture of BSA, Ova and Con. Left and right hand plots display the same data from different angles.

5.2.5.1 Outlier detection and model building of ternary datasets

The same approach was taken with the ternary samples as with the binaries previously discussed. Each samples was prepared in triplicate, spectral outlier rejection was then applied. The total number of spectra taken for outlier rejection including all binaries was 1,998 and the number of spectra replicates rejected was 541. Once rejection of outlying replicate spectra was complete the samples were then screened for outliers in the multivariate space using the same means as described for binary PLS models, 16 out of 685 samples were found to be outlying.

After multivariate outlier rejection was completed samples were split between calibration and validation data sets using the same methods as described for the binary models, split is displayed in Figure 5-17. The number of LVs was chosen for quantification of ternary mixtures using the same method as for the binary models. The number of LVs was compared against RMSEcv. Figure 5-18 shows RMSEcv for BSA, Ova and Con and Table 5-5 shows the actual RMSEcv values and percent changes between LVs. It seems in the ternary PLS model that Ova is more difficult to quantify than BSA and Con, as such the minimisation of RMSEcv will mainly focus on optimisation of Ova estimation. There seems to be a dramatic drop in RMSEcv for Ova at 3 LV but it is significantly higher than RMSEcv BSA and Con. There seems to be significant improvement for Ova at 9 LVs bringing it closer to BSA. However, there is also a significant improvement for BSA estimation at 11 LVs so that was taken as the number of LVs for the model. RMSEcv for Con seems relatively stable beyond 5 LVs. When comparing RMSEcv at 11 LV for the ternary to the BSA-Ova binary at 2 LV they seem comparable.

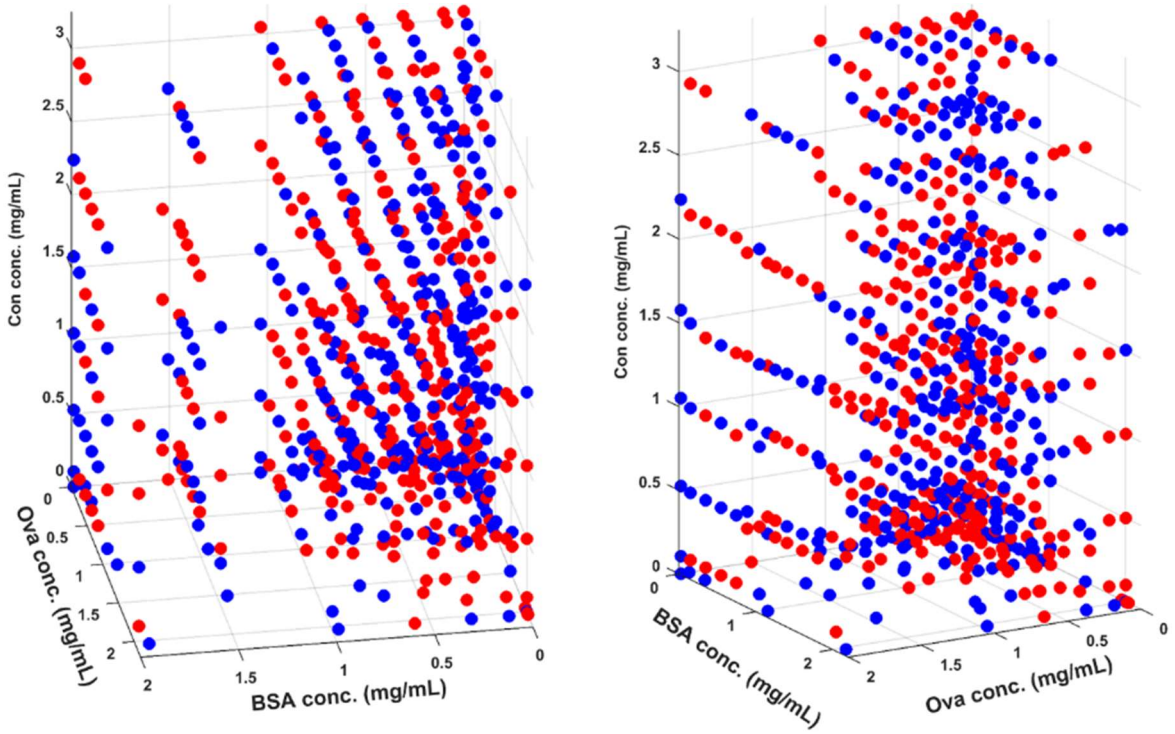

Figure 5-17 Sample split between samples for calibration and validation for ternary and binary PLS model, pH 9. Calibration samples are displayed in blue and validation in red.

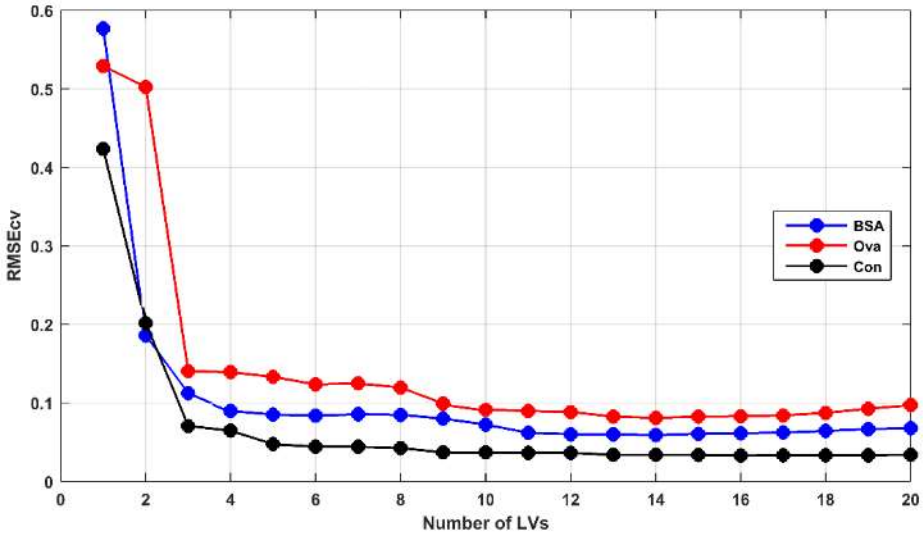

Figure 5-18 RMSEcv for BSA, Ova and Con ternary PLS model building at pH 9.

	RMSEcv (mg/mL)			*Improvement from 1 LV*			*Improvement from previous number of LVs*		
LV number	**BSA**	**Ova**	**Con**	**BSA**	**Ova**	**Con**	**BSA**	**Ova**	**Con**
1	0.577	0.529	0.424	0%	0%	0%			
2	0.186	0.502	0.201	-68%	-5%	-52%	-68%	-5%	-52%
3	0.112	0.140	0.071	-81%	-73%	-83%	-40%	-72%	-65%
4	0.090	0.139	0.065	-84%	-74%	-85%	-20%	-1%	-9%
5	0.085	0.133	0.047	-85%	-75%	-89%	-5%	-4%	-27%
6	0.084	0.123	0.044	-85%	-77%	-90%	-2%	-7%	-7%
7	0.085	0.125	0.044	-85%	-76%	-90%	2%	1%	0%
8	0.085	0.119	0.043	-85%	-77%	-90%	-1%	-4%	-4%
9	0.080	0.099	0.037	-86%	-81%	-91%	-6%	-17%	-13%
10	0.072	0.091	0.037	-87%	-83%	-91%	-9%	-8%	0%
11	0.062	0.090	0.036	-89%	-83%	-91%	-14%	-2%	-1%
12	0.060	0.088	0.037	-90%	-83%	-91%	-3%	-2%	0%
13	0.060	0.083	0.034	-90%	-84%	-92%	0%	-7%	-8%
14	0.059	0.081	0.034	-90%	-85%	-92%	-2%	-2%	0%
15	0.061	0.083	0.034	-89%	-84%	-92%	3%	2%	-1%
16	0.061	0.083	0.033	-89%	-84%	-92%	1%	0%	-2%
17	0.062	0.084	0.033	-89%	-84%	-92%	2%	1%	1%
18	0.064	0.087	0.033	-89%	-83%	-92%	3%	4%	0%
19	0.067	0.093	0.033	-88%	-82%	-92%	4%	6%	0%
20	0.068	0.097	0.034	-88%	-82%	-92%	2%	5%	2%

Table 5-5 Showing RMSEcv and percent change in RMSEcv for BSA-Ova-Con ternary PLS model at pH 9. RMSEcv has been highlighted, green colours represent lower RMSEcv and red higher values.

5.2.5.2 Ternary performance at predicting concentration

The performance of the calibration and validation samples of the PLS model are displayed in Figure 5-19, Figure 5-20, Figure 5-21 and Figure 5-22. Figure 5-19A, B and C are parity plots for BSA, Ova and Con estimation respectively. These plots suggest that the model predictions remain linear in this model. There additionally looks to be no/ very little bias in the model as the slopes are almost equal to 1 and the intercepts are near 0.

Plots Figure 5-19D, E and F show the error in prediction for BSA, Ova and Con respectively. Con is the best estimated of the 3 with the error staying within 0.05 mg/mL for some of the plot and mostly staying within 0.1 mg/mL, the next best estimated is BSA which stays mostly within 0.1 mg/mL error. The worst estimated is Ova which stays mostly within 0.2 mg/mL error. The BSA and Ova data sets are homoscedastic; but there is a tendency for the errors in Con to increase with increasing concentration, when considered in terms of relative error (equation 5-3, all concentrations in mg/mL) but this is not problematic when considered in terms of relative error (equation 5-3, all concentrations in mg/mL) as the increased denominator will cause the relative error to be negligible. The error associated in prediction shown in Figure 5-19D, E and F compared with equivalent binary model performance in Figure 5-11C & D, Figure 5-12C & D, Figure 5-14C & D show that the ternary models do not perform as well as the ternary.

Figure 5-20 show that the expected trend that relative error in BSA increases as BSA concentration decreases and the denominator in 5-3 becomes controlling. There isn't much dependency of BSA error in Ova concentration at low levels of Con but as Con increases this dependency increases, the BSA error with respect to Con concentration remain fairly constant throughout. Ova does show the expected trend of increasing relative error with decreasing Ova concentration but only at the Con levels ≥ 1 mg/mL. Ova absolute relative error in estimation doesn't show a strong dependency on BSA concentration until the greatest levels of Con at 3mg/mL where there is some. The patterns for Ova estimation error are not as strong as the other 2 and there is an element of randomness to the absolute accuracy error. Con is the best

estimated of the 3, possibly because it has the strongest increase in absorbance in terms of mg/mL/AU (see Figure 5-3). In Figure 5-22 there is a link with increasing absolute relative error with decrease Con concentration. There is a small dependency of Con error on Ova concentration, increasing Ova concentration at low Con levels shows increased relative error. This however, is not exacerbated by increasing BSA levels.

The errors associated with calibration and validation data sets seems quite comparable across all proteins. This is promising as it suggests the models have not been over fit, if too many LVs had been added to the model it might be expected that the validation dataset would perform worse than the calibration dataset.

Although there is some dependency of concentration error on the level of competing protein the dependency fairly weak, bearing in mind that increased ranges mean increased assayable range for the analytical method. As there is some dependency in error of on component on another it is conceivable that a PLS model with reduced ranges would predict concentrations with less error. As a test an external dataset was applied to the PLS model.

5.2.5.2.1 Ternary performance on external data set

When multicomponent isotherms were studied a feed plate was generated. This feed plate is prepared in the same way as the control samples for the PLS models. Although feed plates are of known concentration they are always analysed and assayed as an experimental control. The feed plate samples are of known concentration but are prepared at greater concentrations than the ternary PLS models can quantify, as such they are diluted before analysis to allow them to fall into the assayable space. Such a feed plate was prepared for a ternary isotherm at pH 9 with 50 mM NaCl.

As a comparison a scaled down ternary PLS model was created where all three components where input in the ranges 0-1 mg/mL. In order to compare the smaller and larger models an external dataset was applied, this external dataset is the feed plate for the ternary isotherm at 50 mM NaCl, samples were prepared in the range

of 1.4 to 10.4 mg/mL for all 3 components, they were subsequently diluted into the assayable space for both the larger and smaller models before being quantified using both models. This is a useful comparison between larger and smaller models because it replicates the application of the method well with the error in the PLS model as well as any dilution error in addition to the confounding of errors built into the test. It is conceivable that the smaller model contains less error. However, the larger dilution factor required to access the small model space results in the smaller error being multiplied by a larger dilution factor resulting in more error overall in comparison to the larger model. Figure 5-23 suggests the global models are associated with less error overall. It was decided that these larger models had acceptable levels of dependency and the models were taken for use quantifying ternary isotherms. The assayed levels are displayed in Figure 5-24. Many of these samples fall into the “gaps” left in the ternary isotherm control samples, as such they can serve as an additional confirmation that PLS models are fit for purpose in these less populated regions.

Figure 5-25 shows the PLS performance for the PLS control solutions as well as the external data set (pH 9 50 mM feed plate). The external data set behaves similarly to the control samples, the bias in Figure 5-25 A, B and C are similar to Figure 5-19. The error in plots Figure 5-25D, E and F also look similar across the control solutions and the external data set. All of this together suggests that the PLS ternary model works well in the less populated spaced between points.

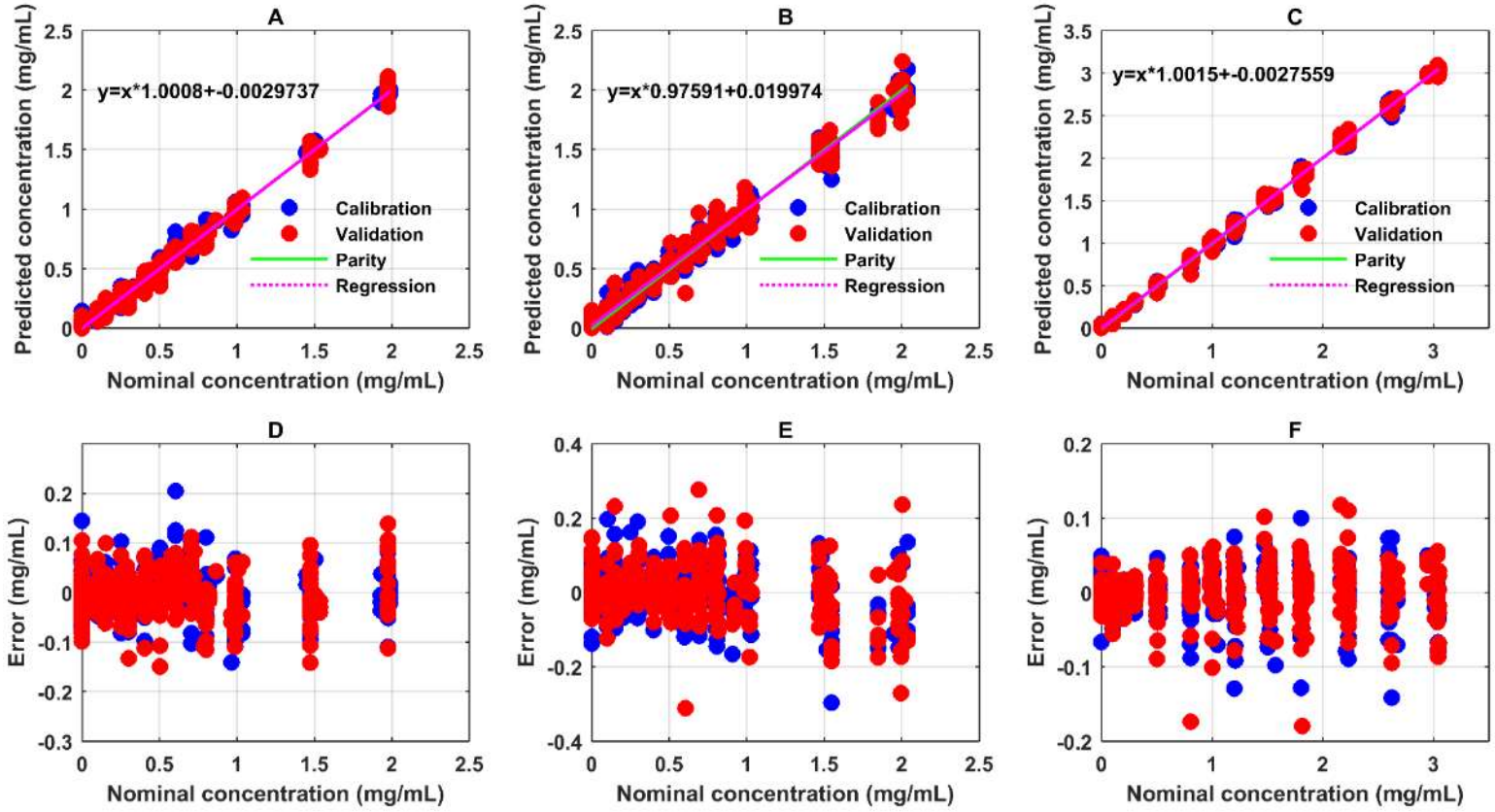

Figure 5-19 Performance of PLS models predicting concentration of individual components in ternary mixtures of BSA-Ova-Con at pH 9. Panels A & D refer to the prediction of BSA, panels B and E refer to prediction of Ova and panels C and F refer to prediction of Con in the ternary mixture. A, B C are parity plots of nominal vs predicted concentration, a linear regression has been fit to the data to assess bias/linearity. Plots D, E and F assess the error in prediction assessing general performance.

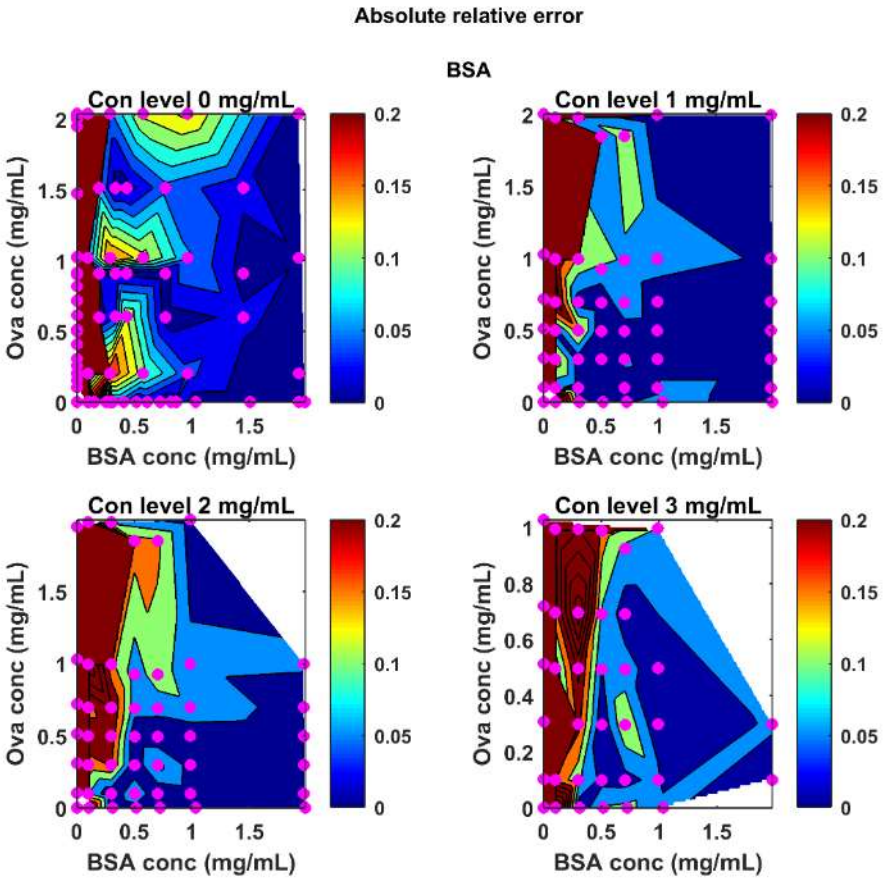

Figure 5-20 Contour plot showing absolute relative error in BSA estimation at different levels of BSA, Ova and Con. Each contour plot shows continuous changes in BSA and Ova and between contours shows the error at different levels of Con. The colour bar has been set from 0-0.2 for every plot for easy comparison. Division by 0 creates values of infinity which are displayed in dark red, relative error values < 0.2 are also displayed as dark red. Calibration and validation points which have been used to create the contours are in pink.

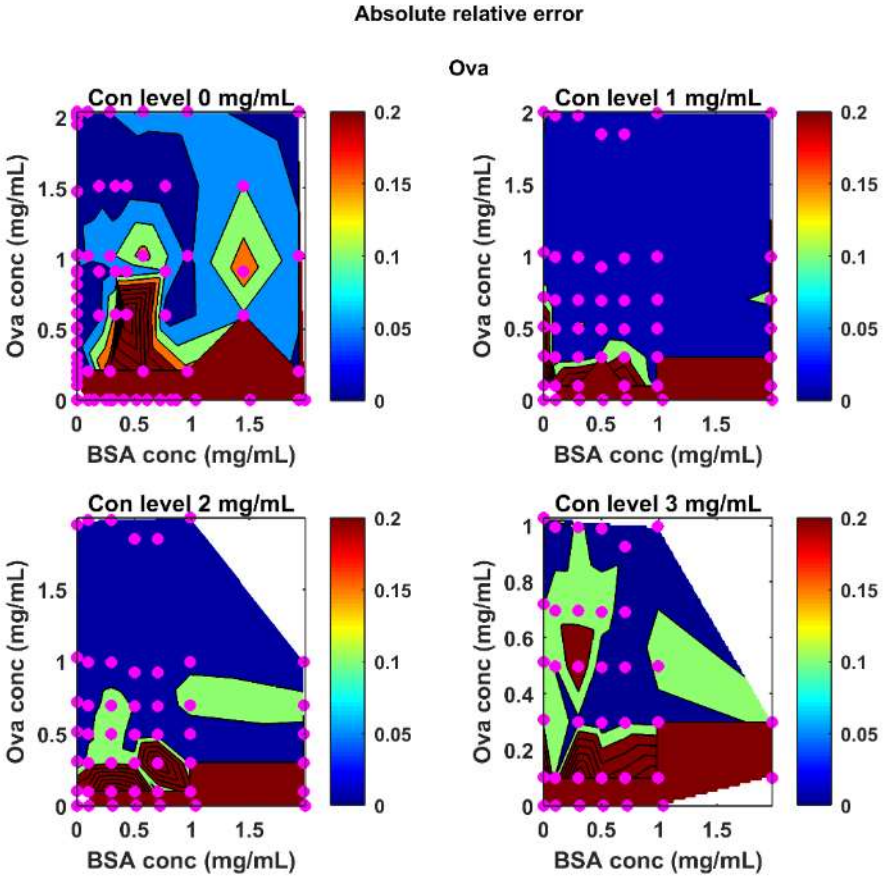

Figure 5-21 Contour plot showing absolute relative error in Ova estimation at different levels of BSA, Ova and Con. Each contour plot shows continuous changes in BSA and Ova and between contours shows the error at different levels of Con. The colour bar has been set from 0-0.2 for every plot for easy comparison. Division by 0 creates values of infinity which are displayed in dark red, relative error values < 0.2 are also displayed as dark red. Calibration and validation points which have been used to create the contours are in pink.

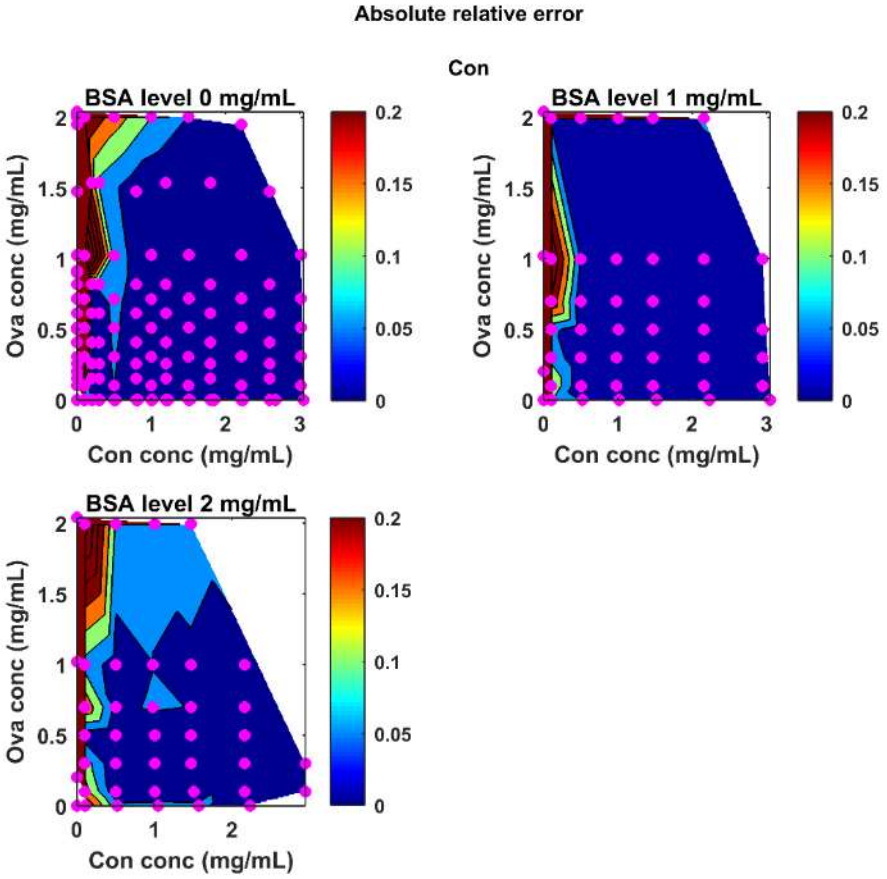

Figure 5-22 Contour plot showing absolute relative error in Con estimation at different levels of BSA, Ova and Con. Each contour plot shows continuous changes in Con and Ova and between contours shows the error at different levels of BSA. The colour bar has been set from 0-0.2 for every plot for easy comparison. Division by 0 creates values of infinity which are displayed in dark red, relative error values < 0.2 are also displayed as dark red. Calibration and validation points which have been used to create the contours are in pink.

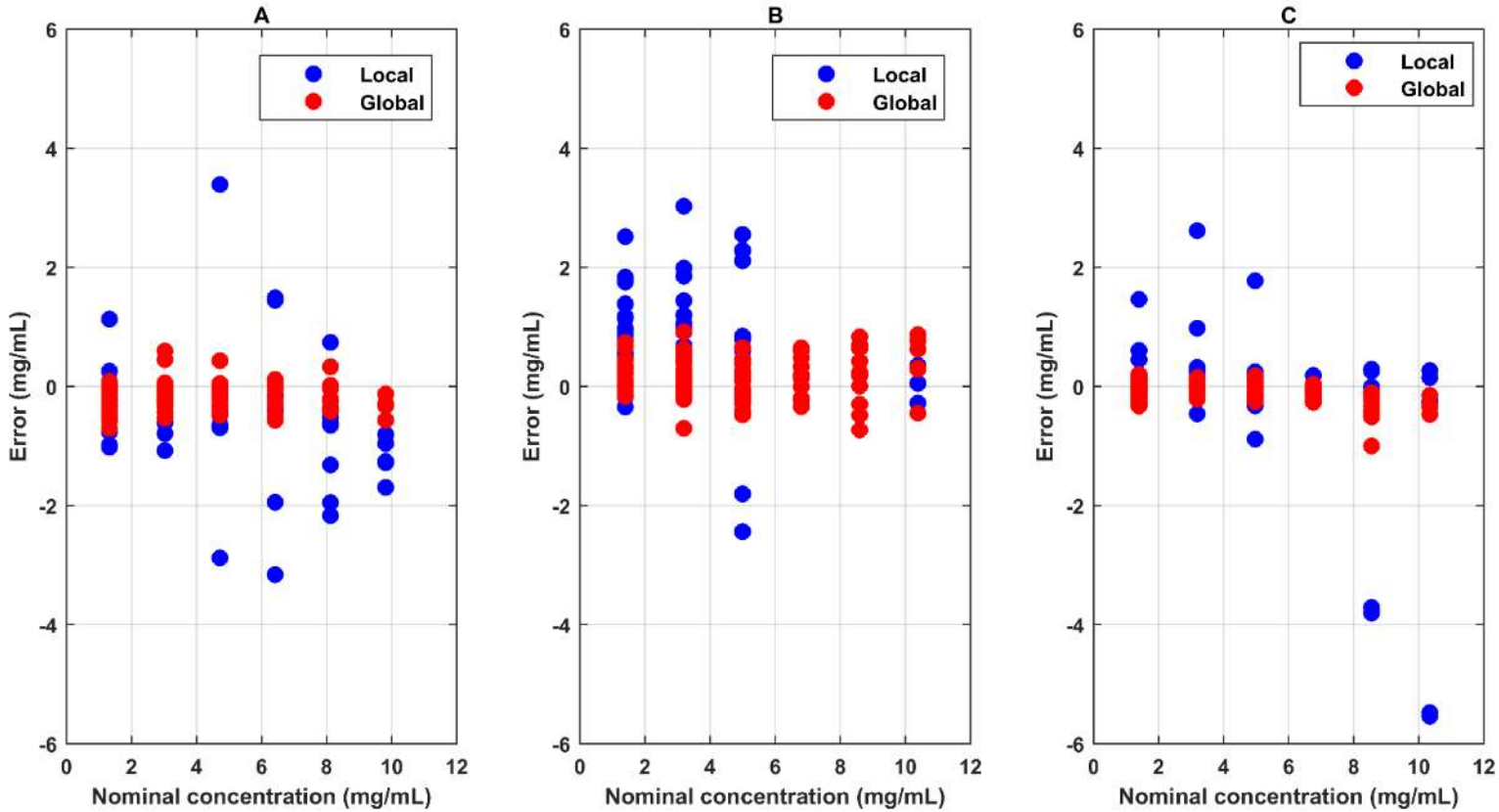

Figure 5-23 Comparison of local and global ternary PLS models for the error in prediction of BSA, Ova and Con concentrations in panel A, B and C respectively. Local models ran from the range 0-1 for all 3 components, global models ran from 0-2 for BSA & Ova and 0-3 for Con, samples had different dilution factors applied in order for them to fall into the assayable space.

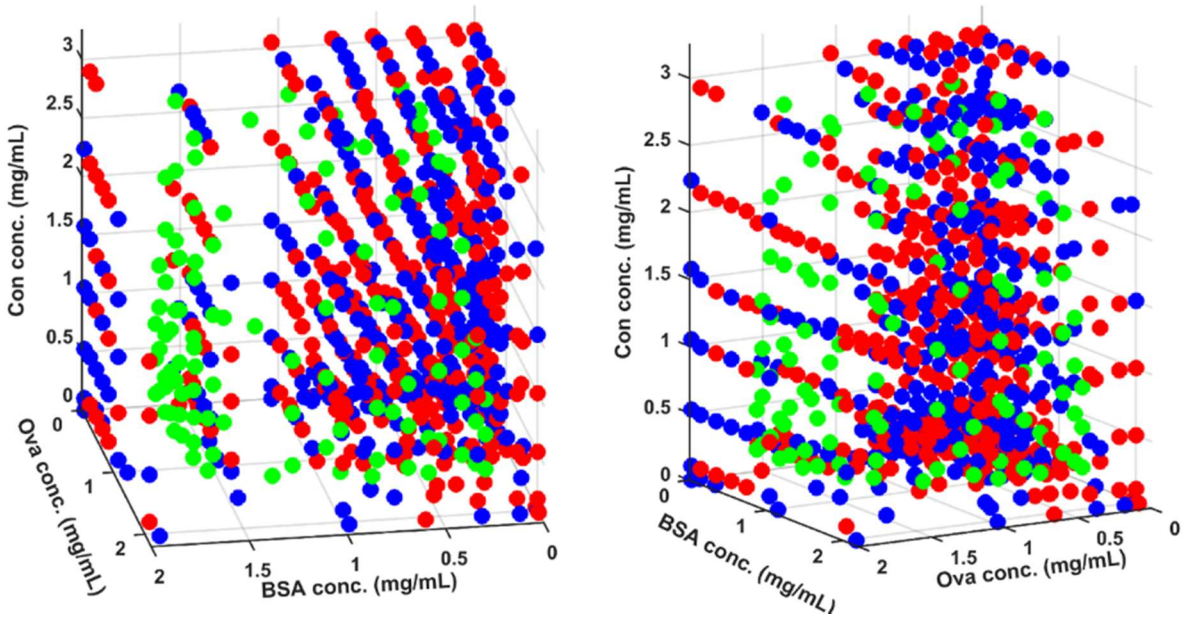

Figure 5-24 All control samples for ternary PLS model building, pH 9. Calibration samples are displayed in blue and validation in red, external sample feed solutions for 50 mM NaCl have been included in green.

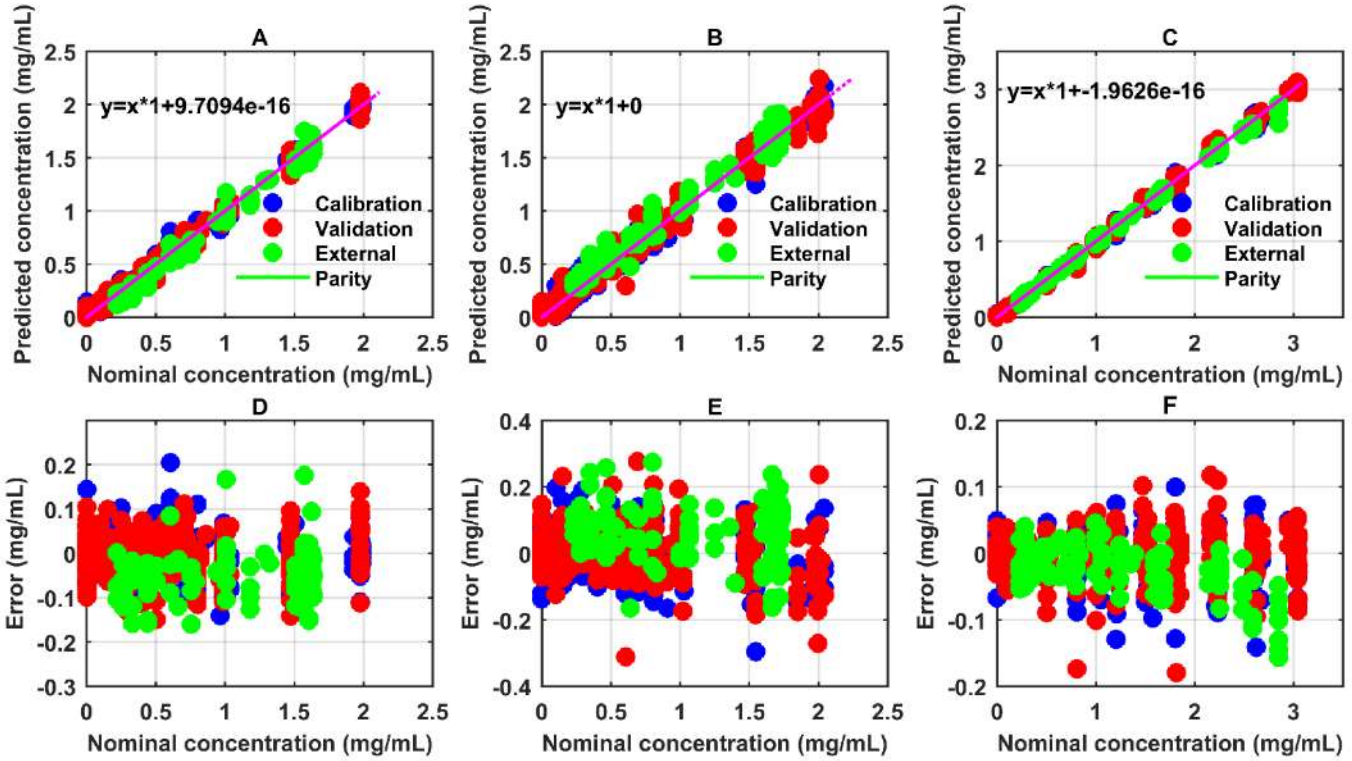

Figure 5-25 Performance of PLS models predicting concentration of individual components in ternary mixtures of BSA-Ova-Con at pH 9. Panels A & D refer to the prediction of BSA, panels B and E refer to prediction of Ova and panels C and F refer to prediction of Con in the ternary mixture. A, B C are parity plots of nominal vs predicted concentration, a linear regression has been fit to the data to assess bias/linearity. Plots D, E and F assess the error in prediction assessing general performance. An external data set has been added to the calibration and validation data set, this external data set is the feed plate estimations for a ternary isotherm in 50 mM NaCl pH 9.

5.3 Discussion and summary

It was discussed in section 2.5 HPLC separation of model mixture that analytical chromatographic separation of the BSA, Ova, Con model mixtures would be challenging. No single analytical method would likely allow for baseline separation of all 3 proteins from one another. One solution would be to run multiple separations, assaying BSA using HIC, fractionating the co-eluting Ova, Con peak and separating that on SEC. This would be prohibitively slow when studying multicomponent isotherms as the number of samples involved is high. Alternatively a single analytical separation could be performing using AEX and the overlapping peak could have a deconvolution method applied. This is a more feasible method but would still hinder throughput when studying multicomponent isotherms. The alternative solution presented here is to use UV spectra in conjunction with multivariate data analysis (PLS) in order to assay individual protein concentrations within a mixture.

This method has been applied for the quantification of proteins within mixtures before, it has also been used in the application of multicomponent isotherms but with a caveat. In the previous application of this method to binary isotherms one of the proteins contained a heme group which facilitates the quantification of proteins in this way. As most proteins do not contain prosthetic groups offering additional chromophores the work presented here is a useful extension of that branch of investigation. Methods to clean up the data before commencing model building were also presented. First the univariate spectra data had outliers rejected using median absolute deviation. Additionally outliers in multivariate space were also identified and removed before model building.

Further advances made in this chapter were to increase the assayable space for the PLS models which allows for increased sensitivity for samples which would otherwise require dilution. Moreover, the PLS models presented here have been heavily over populated using a space filling grid layout. As well as increasing the robustness of the PLS models by making them less dependent on fewer samples it also offers the possibility of monitoring the performance of the PLS models at regular discrete points and monitoring

the dependency of error of protein A on the concentration of protein B, which has not been discussed before. The use of local models which span part of the range of the global model has also not been presented before in this application. It can be useful in reducing error in regions where the signal of protein being assayed in lower in comparison to the total signal.

The binary PLS models presented performed well and will be taken forward for use in quantifying binary isotherms, the only caveat being that some dependence of the error on one component was found in the concentration of the other component. It was decided that the slight dependence observed was an acceptable trade-off for increased assayable space. Ternary PLS models did not perform as well binary models with error generally being greater, in the ternary PLS model Ova was the worst performer. Decreased models size running from 0-1 mg/mL for all components actually performed worse at quantifying an external data set. This is because when using smaller models samples require larger dilution factors meaning any error in the model is also multiplied by the larger dilution factor. Moreover, there is also more error associated with larger dilutions. The ternary models are also taken forward for the quantification of multicomponent isotherms.

Ternary PLS models for the quantification of proteins have been demonstrated previously by other groups (Hansen et al., 2013), the findings made are in agreement with what was shown in this thesis. In agreement with this work they also found decreased model performance when going from binary solutions to ternary as well as the need to increase in the number of LVs when studying more complex mixtures. As part of their work they took the spectra of 26 proteins, including BSA, Ova and Con and assessed their similarity. Figure 5-26 shows a principle component analysis (PCA) of the protein spectra for the first 2 principle components (principle components in a PCA model are analogous to LVs in a PLS model). PCA is a similar method to PLS except it only analyses and compares X data (spectra) and makes no attempt link it to Y data (concentrations).

Visual inspection of Figure 5-26 shows that BSA, Ova and Con lie relatively close to one another; one of the findings when comparing these model spectra was that similar

spectra are more difficult to quantify. Moreover, they suggest that proteins in the red and pink clusters are associated with greater absorbance and better estimations of their concentration. This is also in agreement with the finding made in this thesis that Con (belonging to the pink cluster) is always the most accurately assayed protein in all mixtures. Furthermore, work done for this thesis but not presented showed haemoglobin (Hb) (red cluster) was also well quantified in all mixtures studied which is in agreement with this group's findings. It was also demonstrated that with improved equipment such a dual beam spectrophotometer the model performance can be improved significantly for these higher order models (Hansen et al., 2013). However, the use of classic cuvette and spectrophotometer would not be a realistic possibility for the application studied here as the sample volume from a slurry plate would be too small for such analysis. Additionally, the HT nature of the analytical method would be lost in order to gain better model performance.

The use of UV spectra in conjunction with multivariate data analysis offers a powerful alternative for rapid quantification of protein mixtures. In the following chapters it will be applied for the quantification of 3 binary and 1 ternary protein mixture and a range of pHs and NaCl concentrations.

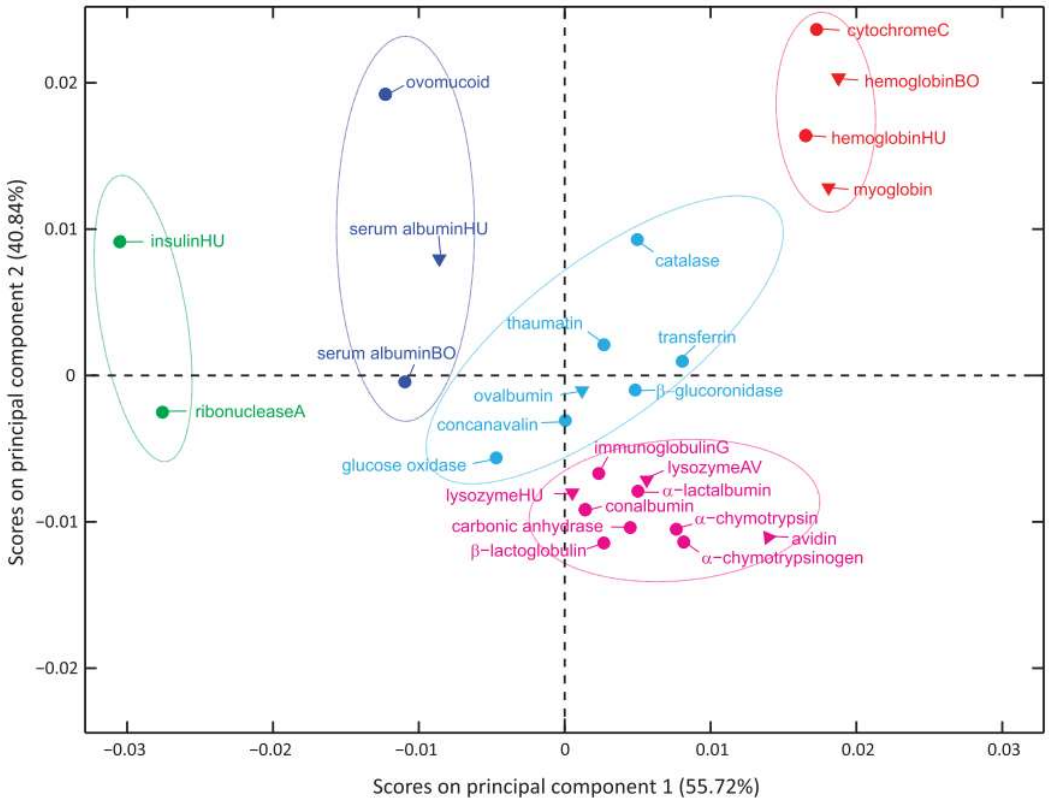

Figure 5-26 PCA scatter plot of spectral data for 26 commonly studied proteins. Colours represent different clusters assigned to the proteins, triangles represent proteins taken forward for PLS model building (not relevant to this discussion). Reproduced from (Hansen et al., 2013).

Chapter 6 Binary component isotherms and their associated error

6.1 Aim and objectives of chapter

In Chapter 4 a platform methodology was developed and presented for the generation of single component adsorption isotherms using high-throughput (HT) methods and automated liquid handling systems (ALHS). In Chapter 4 quantification of adsorbed concentration in isotherms was explored using 2 methods: Firstly the traditional method; here the starting liquid concentration is known and the ending concentration is assayed, the adsorbed concentration is then calculated via mass balance and is referred to as q. The alternative method is to elute what was adsorbed to the resin and assay it creating an orthogonal measure of adsorbed concentration referred to as q^*. In chapter Chapter 5 a rapid analytical method for the quantification of individual protein titres from binary and ternary mixtures using UV spectra and multivariate data analysis was presented and discussed.

The aims of this chapter is to present binary adsorption isotherms for BSA-ovalbumin (Ova) on the strong anion exchanger Capto^{TM} Q at two pHs and a range of NaCl concentrations. BSA-conalbumin (Con) and Ova-Con mixtures using the same chromatography system were also studied, for brevity these datasets as well as some others are displayed in Appendix A: Additional binary isotherm datasets. Additionally the propagation of both systematic and random errors in q and q^* for these multicomponent isotherms will be explored. The specific aims of the chapter are:

- Present the following multicomponent isotherms using both q and q^* measures of adsorbed concentration:
    - BSA-Ova: pH 9; 4 NaCl levels

The following conditions are presented in Appendix A: Additional binary isotherm datasets:

- BSA-Ova pH 8; 3 NaCl levels

    - BSA-Con: pH 9; 3 NaCl levels, pH 8 1 NaCl level
    - Ova-Con: pH 9; 3 NaCl levels, pH 8 1 NaCl level
- Present the propagation of both systematic and random errors in adsorbed concentration (both *q* and *q**).
    - Systematic errors are introduced by deviations in accuracy of the analytical method.
        - Chapter 4 dealt with the minimisation of systematic errors associated with slurry plate and ALHS usage. As such the contribution of systematic error from these sources has been taken as negligible.
    - Random errors are introduced from a variety of sources during experimentation.
- The potential effect of error on *q* and *q** measures of adsorbed concentration as well as a discussion of which measure is best suited to what condition. Binary isotherms and their error propagation

Binary and ternary isotherms were generated using the method discussed previously in Chapter 4. The HT slurry plate chromatography format known as the PreDictor™ plate was used. Every micro-well in the slurry plated contained 20 µL of Capto™ Q resin and was deployed in conjunction with an ALHS to facilitate isotherm study. The ALHS was optimised to handle proteinaceous solutions with maximal accuracy as discussed in Chapter 4. In order to assay individual protein concentrations in a mixture, UV spectra were recorded and analysed using multivariate models which calculate the solution composition from the spectra as discussed in Chapter 5. These models use the multivariate method known as partial least squares regressions (PLS) to link variation in X (spectra) to variation in Y (titres).

In Chapter 4 systematic error caused by experimental choices such as phase ratio as well as liquid handling accuracy on the ALHS was minimised. It was also found that quantification of adsorbed concentration in isotherms by elution can reduce random variation in specific situations. The sources of systematic and random error for this platform methodology have been studied and reduced in previous chapters.

In this chapter the high throughput protein quantification method is applied to the platform in order to study binary isotherms. Since experimental systematic error has been minimised for the platform method, the remaining systematic error associated with protein quantification is propagated here as one assessment of the effectiveness of the method.

The random error remaining in the platform methodology as well as the random error associated with quantification of mixtures is propagated through the binary isotherms for both q and q^* measures of adsorbed concentration. A detailed account of the error propagation is given in Chapter 3.

The samples generated in order to build PLS methods discussed in Chapter 5 were well populated and the errors at any given point in the assayable space were quantified. Using this map of accuracy errors in conjunction with the Matlab function "griddata" which interpolates values between points, an estimate of systematic error at any concentration in the assayable space can be obtained. This systematic error in assayed concentration can then be propagated to a systematic error in q and q^*.

For the q estimate in adsorbed concentration the only source of systematic error in the estimation of equilibrium concentration as shown in equation 3-5. For the q^* estimate in adsorbed concentration elution, as calculated in equation 3-7, there are more sources of systematic error: All the wash/elution fractions analysed using the PLS method can introduce systematic error: That systematic error can be either positive or negative, increasing or decreased the estimated concentration. The systematic errors in q^* calculation were allowed to cancel. Meaning that if elution fraction 1 was associated with an over estimate and fraction 2 an underestimate these errors were allowed to cancel one another out.

As previously mentioned there are multiple sources of random error. An estimate of these errors has been gleaned either from in house experimental data or taken from the literature and the equations used to propagate the error are described in Table 6-1 which were used in conjunction with equations 3-5 and 3-7.

$$q = (C_0 - C_{eq}) \times F \qquad 3\text{-}5$$

$$q^* = \frac{(\sum_{i=1}^{n} m_{w,i} - m_{rw,i} + \sum_{i=1}^{n} m_{e,i} - \sum_{i=1}^{n-1} m_{re,i} - m_{req})}{V_{resin}} \qquad 3\text{-}7$$

Calculation type	*Example*	*Random error*
Addition or subtraction	$x = a + b - c$ *(6-1)*	$\sigma_x = \sigma_a + \sigma_b + \sigma_c$ *(6-2)*
Multiplication or division	$x = \frac{a \times b}{c}$ *(6-3)*	$\frac{\sigma_x}{x} = \frac{\sigma_a}{a} + \frac{\sigma_b}{b} + \frac{\sigma_c}{c}$ *(6-4)*

Table 6-1 Propagation of maximum random error. σ represents random error associated with value. These equations are different to those used to propagate standard deviation as these propagate maximum random error (The American Society of Mechanical Engineers, 2005).

6.1.1 Modelling random error in liquid concentration estimation

The estimation of concentration using UV spectra and PLS models is associated with some random error. In order to propagate the random error in estimation, a relationship between concentration estimate and random error needs to be established. The control samples used to calibrate and validate the PLS models were prepared in triplicate and then averaged before model building. An example set of spectra are displayed in Figure 6-1 with the average percent relative standard deviation (%RSD) of the spectra and the %RSD in PLS estimation of that spectra. In order to estimate the random error in the system, the triplicate spectra were supplied to the PLS models in order to generate estimates for each individual spectra, from which the standard deviation and average estimate were calculated. This random error was then fitted to the average concentration estimates using linear regression.

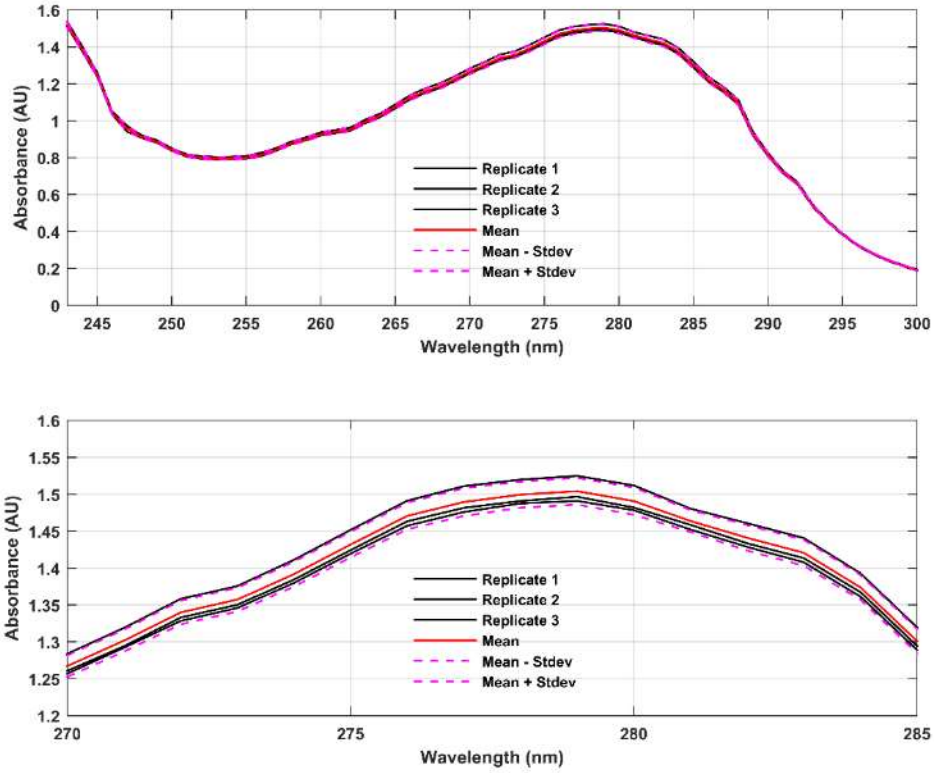

Figure 6-1 Blank corrected spectra of BSA-Ova binary mixtures at pH 9. The top plot displays the whole spectrum from 243-300 nm and the bottom spectrum displays the same data but enlarged. Plots display 3 spectral replicates as well as the mean and the mean +/- the standard deviation. The nominal concentrations of BSA and Ova were 3.9 and 1.0 mg/mL respectively and the average predicted concentration of BSA and Ova was 3.8 and 1.0 mg/mL respectively. The average percent relative standard deviation (%RSD) of the spectra is 1% and the %RSD in BSA and Ova prediction is 2% and 6% respectively.

6.1.1.1 Modelling random error in concentration estimation

6.1.1.1.1 BSA-Ova

A simple linear regression was applied linking average concentration estimates to standard deviation of triplicate estimations. BSA standard deviation and Ova standard deviation were modelled separately and no dependency of BSA standard on Ova concentration was found. Conversely, no dependency of Ova standard deviation on BSA concentration was found. In all cases the p-values for associated coefficients were greater than 0.05. Linear regression of random error is displayed in Figure 6-2.

There is significant variation in the standard deviation when which makes fitting a line of regression to the concentration estimate and standard deviation challenging. The statistically significant p-values (all <0.05) suggests there is a relationship between concentration and standard deviation meaning modelling the noise in this way is the more appropriate than simply taking a fixed average for all concentration levels.

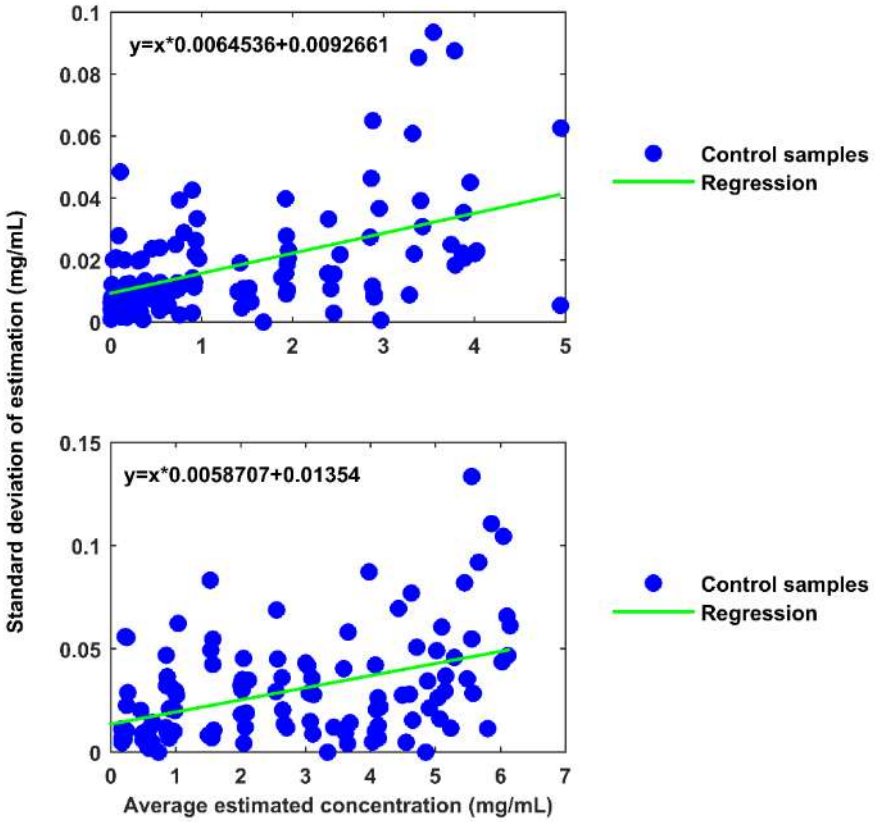

Figure 6-2 Modelling standard deviation of triplicate estimates against average estimate. Top panel displays modelling BSA random error and bottom panel Ova random error, lines of regression and parameters are displayed in the top left of each plot.

6.1.1.1.2 BSA-Con

A simple linear regression of BSA standard deviation dependant on average estimated BSA concentration as well as average estimated Con concentration was performed and is displayed in Figure 6-3. The parameter pertaining to both the estimated BSA and Con concentrations had p-values <0.05 and so a linear regression dependant on both concentration estimates was taken forward.

6.1.1.1.3 Ova-Con

A simple linear regression of Ova standard deviation dependant on average estimated Ova concentration and average estimated Con concentration was performed and is displayed in Figure 6-4. The parameter pertaining to both the estimated Ova and Con concentrations had p-values <0.05 and so linear regression dependant on both concentrations executed. The intercept modelling Ova random error had a p-value greater than 0.05 and so was set to 0.

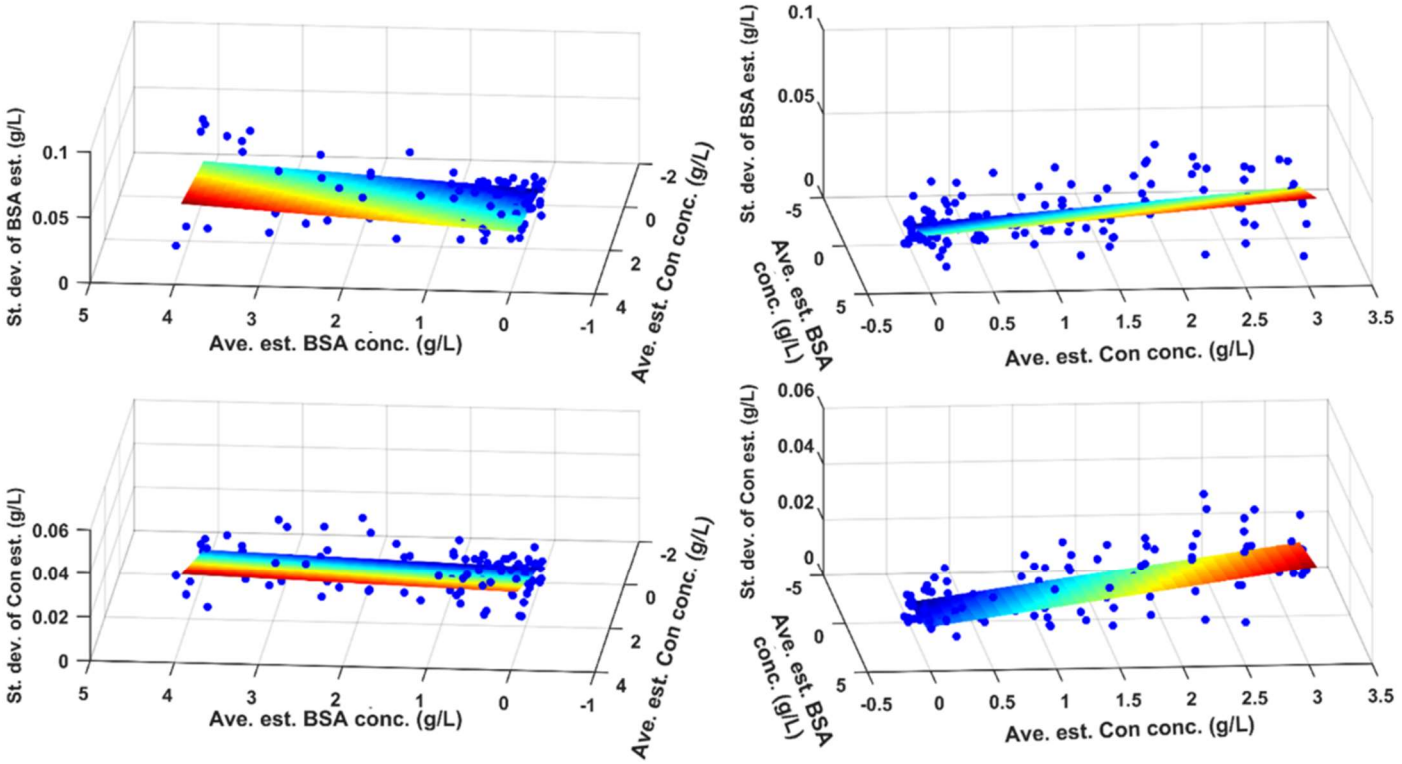

Figure 6-3 Modelling standard deviation of triplicate estimates against average estimate for BSA-Con binary mixtures. Top 2 panels displays modelling BSA random error from perpendicular angles and bottom 2 panels Con random error from perpendicular angles.

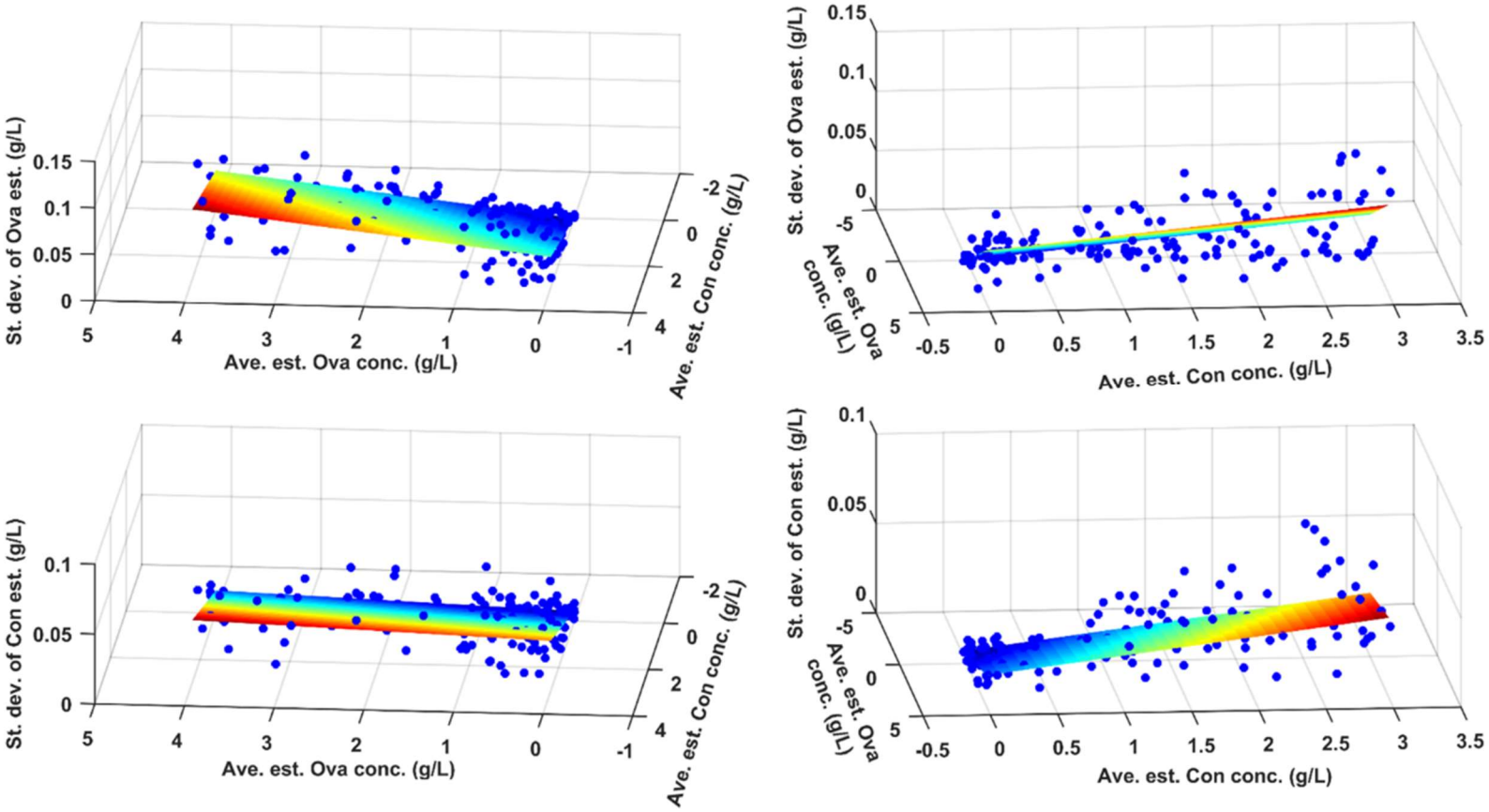

Figure 6-4 Modelling standard deviation of triplicate estimates against average estimate for Ova-Con binary mixtures. Top 2 panels displays modelling Ova random error from perpendicular angles and bottom 2 panels Con random error from perpendicular angles.

6.1.2 BSA-Ova binary pH 9 isotherms and their error propagation

6.1.2.1 0 mM NaCl isotherm

The binary BSA-Ova isotherm at pH 9 in 0 mM NaCl is presented in Figure 6-5 along with agreement between adsorbed concentrations calculated at equilibrium (q) and via elution (q^*). The shape of the BSA and Ova isotherms in Figure 6-5A and B in the presence of little competition retain a similar shape to what was observed for the single component isotherms in the same buffer conditions as shown in Figure 4-11. BSA keeps a highly rectangular appearance with very little increase in liquid concentration with increasing adsorbed concentration until the plateau is reached. As in the single component isotherm (SCI) Ova has decreased rectangularity in comparison to BSA, there are some increases in liquid concentration with increasing adsorbed concentration before the plateau is reached in the non-competitive region of the isotherm.

The isotherms show significant competition and displacement for both components. BSA which is the most retained species is effectively displaced by small quantities of Ova with 3 mg/mL starting concentration of Ova deceasing the plateau and decreasing the rectangularity of the isotherm. Displacement continues until a starting concentration of Ova at 9 mg/mL where additional Ova has little additional effect. Whilst Ova has a lower saturation capacity than BSA and is the less retained species BSA is not as effective at displacing Ova as Ova is at displacing BSA. At 3 mg/mL BSA has had very little effect on Ova whereas Ova has displaced BSA significantly at this level. BSA does reduce the saturation capacity of Ova and decreases the initial slope of the isotherm as well as soften the corner with increasing BSA starting concentration. There again seems to be little additional effect of displacement above 9 mg/mL starting concentration BSA.

Figure 6-5C and D show very good agreement between q and q^*. This is significant indication that the methodology is working well. Systematic overestimation in assayed liquid concentrations would lead to systematic

underestimation of adsorbed concentration using the equilibrium mass balance calculation 3-3 and systematic overestimation in the adsorbed concentration using the elution fractions 3-7. The fact that these two orthogonal methods of measurement agree so closely suggests this data set is free of significant systematic error.

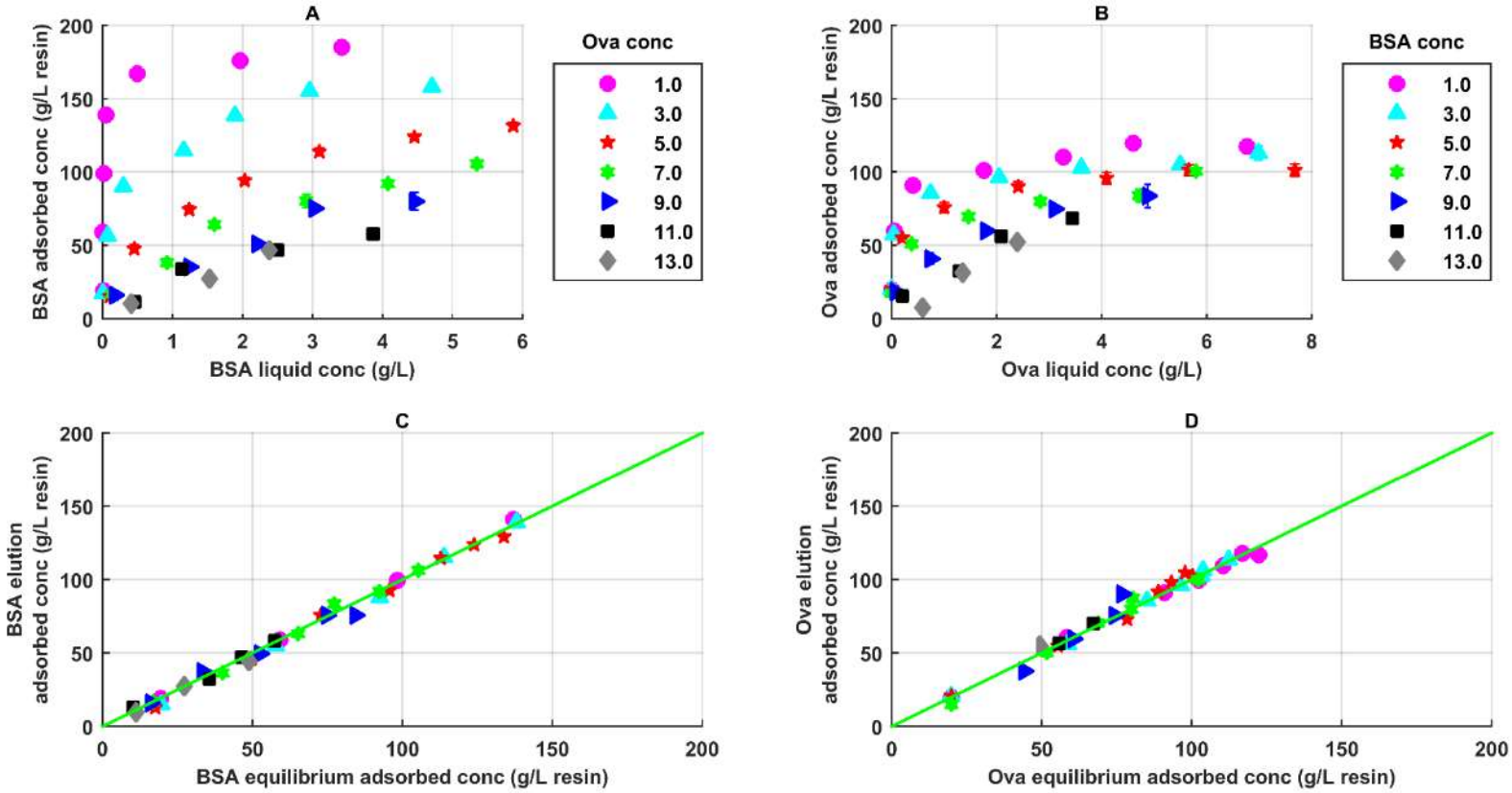

Figure 6-5 BSA-Ova binary isotherm pH 9 0 mM NaCl. Panels A & C present BSA liquid and adsorbed concentrations, B & D represent Ova data. Panels A and B present the binary isotherms at different starting concentrations of the competitor displayed in the figure legend in mg/mL, the liquid concentration is the average of q and q after outlier rejection. Error bars in panels A and B represent experimental standard deviation. Panels C and D present agreement between q and q*with a line of parity in green.*

6.1.2.2 0 mM error propagation

Figure 6-6 shows the BSA-Ova binary isotherm at pH 9 0 mM NaCl as a contour plot, this has been included as a reference to facilitate comparison to propagated error contour plots. The isotherm contour plot shows the expected trends with adsorbed concentration of BSA increasing with increasing liquid concentration of BSA. The equivalent trends can also be seen for the Ova part of the isotherm. The displacement of BSA with increasing equilibrium liquid Ova concentration can be seen and the equivalent behaviour is also seen for the Ova data.

*6.1.2.2.1 Systematic error in q and q**

Figure 6-7 shows the propagated systematic error in *q* for the isotherm. The left hand plots show that the propagated error reaches a maximum of 2 mg/mL resin for BSA and 5 mg/mL for Ova and that there is no pattern within the plot. However, there is a pattern across the plots: When the systematic error in mg/mL *q* is high for BSA it is also high for Ova. This is because the PLS model has a tendency to compensate across estimations. This means when one compound is overestimated the other is underestimated. Because only absolute error is displayed in the contour plots which only accounts for the magnitude of the error and not the direction, the areas of larger compensation come at as high values across both components. Both BSA and Ova relative error in *q* (described by equation 6-5) reach a maximum of approximately 0.05 or 5% which is acceptable despite the tendency for compensation.

Figure 6-8 shows the propagated systematic in *q**. Left hand plots show that the propagated error reaches a maximum of 3 mg/mL resin for BSA and 4 mg/mL for Ova and that there is no pattern within the plots. Both BSA and Ova relative error reach a maximum of approximately 0.05 or 5% which is acceptable. Part of the contour plot for systematic error in *q** (Figure 6-8) is missing. This is because some of the elution fraction concentrations were too high and fell outside the assayable space and so have been omitted. They could have been diluted but were not due to an experimental oversight.

*6.1.2.2.2 Random error in q and q**

Figure 6-9 shows the propagated random error in *q*. Left hand plots show that the propagated error reaches a maximum of 12 and 11 mg/mL resin for BSA and Ova respectively. There is a strong pattern of increasing random error with increasing assayed concentration for the protein being presented in the left hand plots. This pattern comes about because random error in g/L increases with assayed concentration and this error directly propagates into the adsorbed concentration as shown in Figure 6-2. There is a second driver behind the random error which is the random error in the starting concentration C_0. The C_0 random error is a constant proportion of the corrected starting concentration meaning at higher C_0 concentrations the error is greater when considered in terms of mg/mL. The random error of the phase ratio *F* as calculated in equation 3-6 is constant throughout the isotherms as the phase ratio is constant. Both BSA and Ova relative error reach a maximum within 0.10 to 0.15 which is acceptable. The high points of relative error are also dominated by the denominator in equation 6-5 as the left hand plots show error in terms of mg/mL resin are low at these points but the adsorbed concentration shown in Figure 6-6 is low at these sample points.

Random error in *q**, displayed in Figure 6-8, shows increases when adsorbed concentration is high, the driver for this has the same root cause as for equilibrium measurement. When the elution fractions contain more protein the random error is larger. The relationship is different from the random error in *q* as elution adsorbed concentration is linked directly to elution 1 concentration which is in turn linked to adsorbed concentration. The random errors in fraction volumes and resin volume in equations 3-6 and 3-8 are constant throughout the binary isotherm.

6.1.2.2.3 Summary of propagated error

Overall the propagation of systematic (accuracy) and random errors (precision) are under good control for both equilibrium an elution for this binary isotherm. There is no obvious pattern in systematic error within the components. However, there is a tendency for *q* systematic error to be higher in mg/mL terms in the same location

across the components in the binary isotherm. Interestingly, the relationship between adsorbed concentration and random error is different for q and q^*.

$$\text{Relative error} = \frac{\text{Propagated error}}{\text{Experimentally observed adsorbed concentration}} \quad \textit{6-5}$$

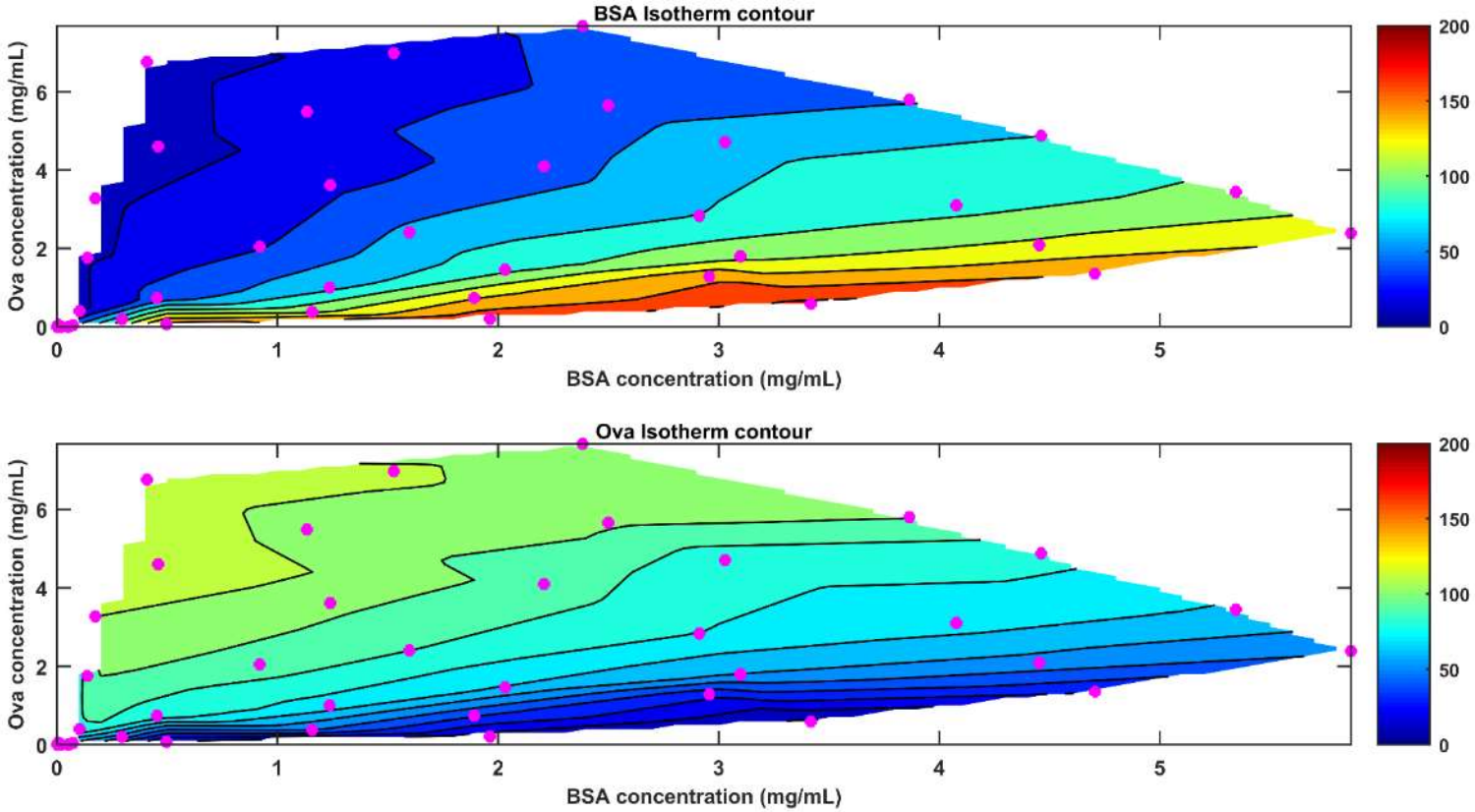

Figure 6-6 Binary isotherm for BSA-Ova pH 9 0 mM NaCl. Isotherm has been drawn as a contour plot as a reference for error propagation plots. x and y axis display average equilibrium concentrations of BSA and Ova measured directly after outlier rejection, contours display adsorbed concentration of BSA in the top panel and Ova in the bottom panel, these have been taken as averages of q and q.*

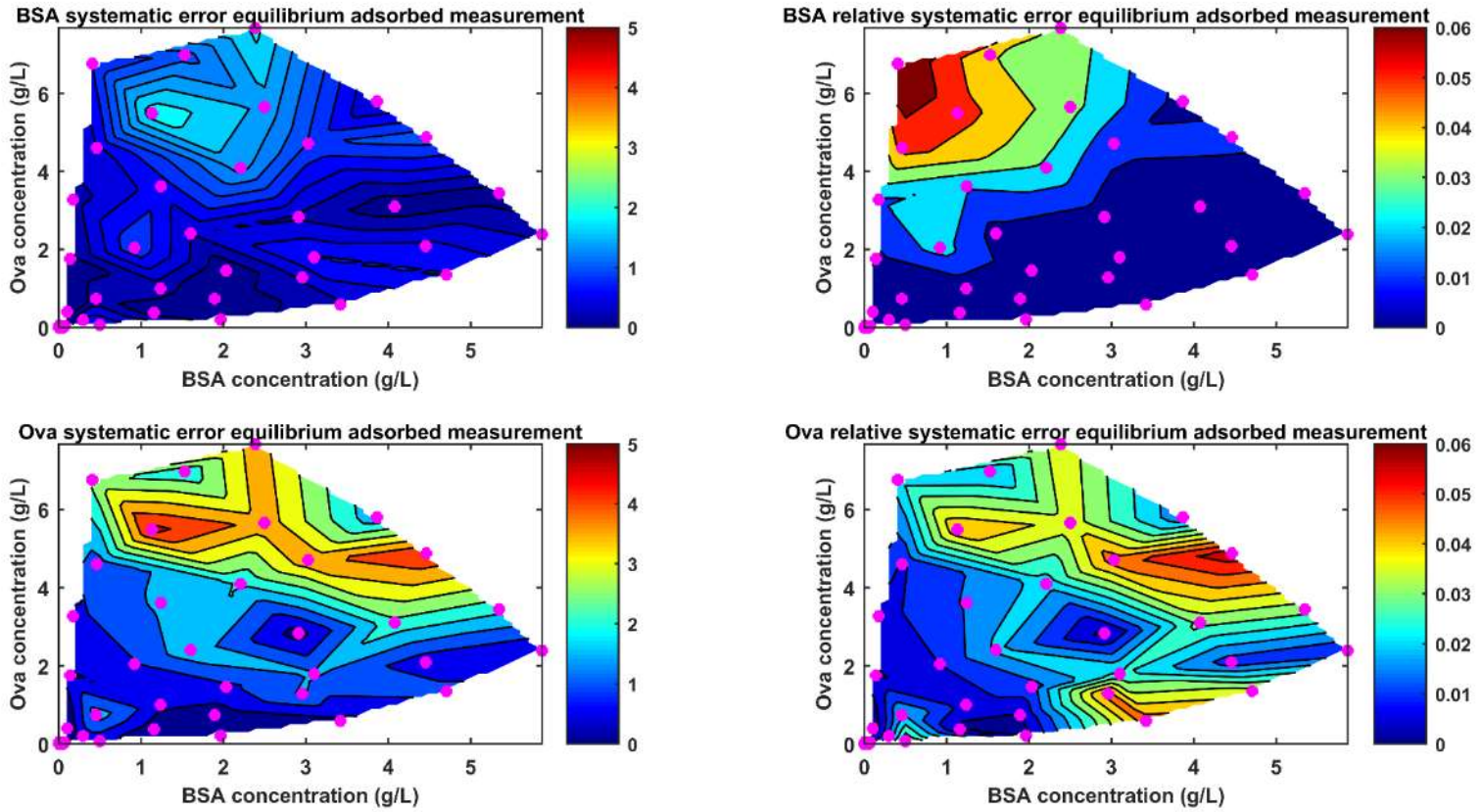

Figure 6-7 Contour plots showing propagated systematic absolute error in adsorbed concentration calculated using q for BSA-Ova binary isotherms at pH 9 0 mM NaCl. Top 2 panels show propagated error for BSA and bottom 2 for Ova. Left hand panels show propagated error in adsorbed concentration mg/mL resin and right hand panels show relative error in adsorbed concentration.

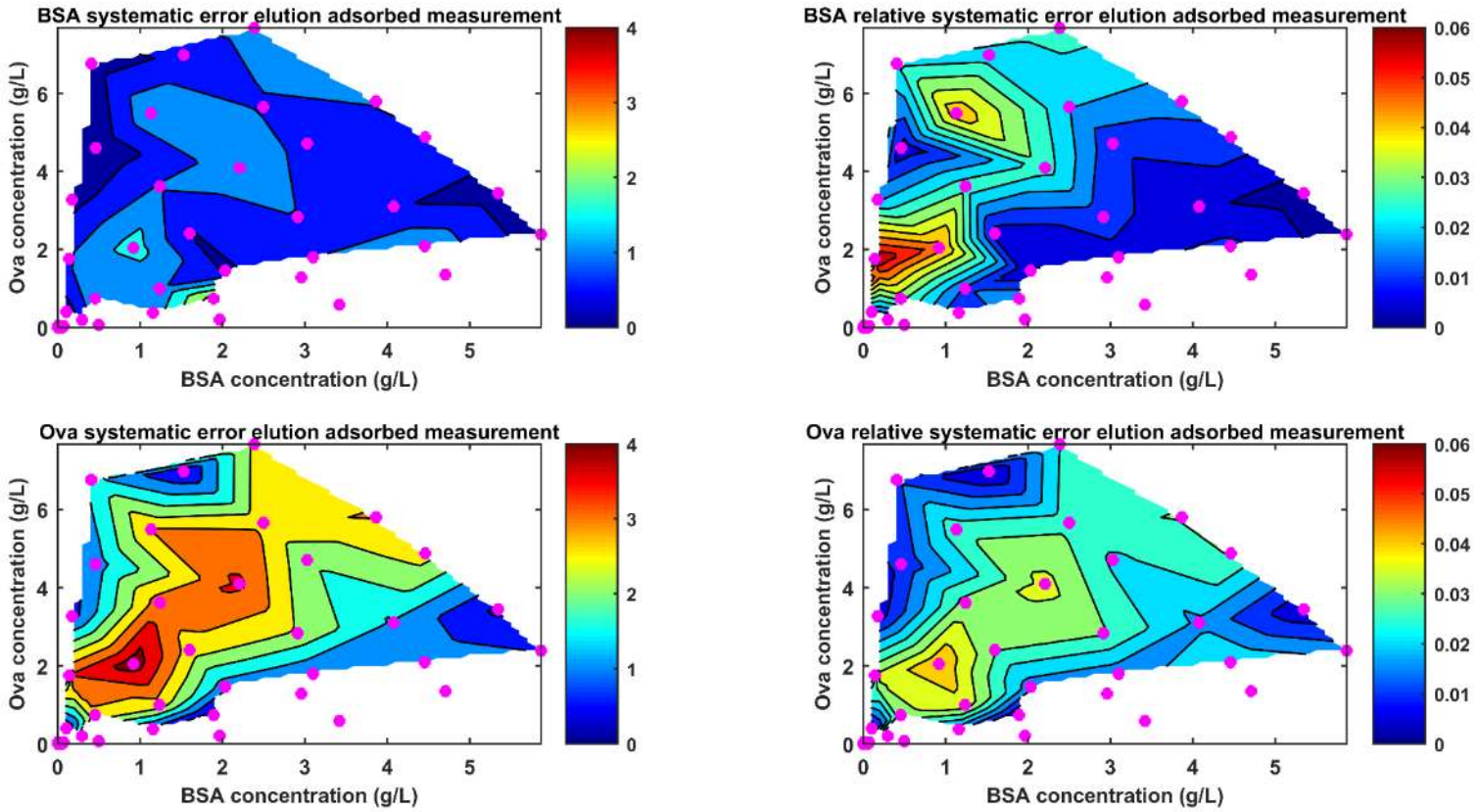

*Figure 6-8 Contour plots showing propagated systematic absolute error in adsorbed concentration calculated using q*for BSA-Ova binary isotherms at pH 9 0 mM NaCl. Top 2 panels show propagated error for BSA and bottom 2 for Ova. Left hand panels show propagated error in adsorbed concentration mg/mL resin and right hand panels show relative error in adsorbed concentration.*

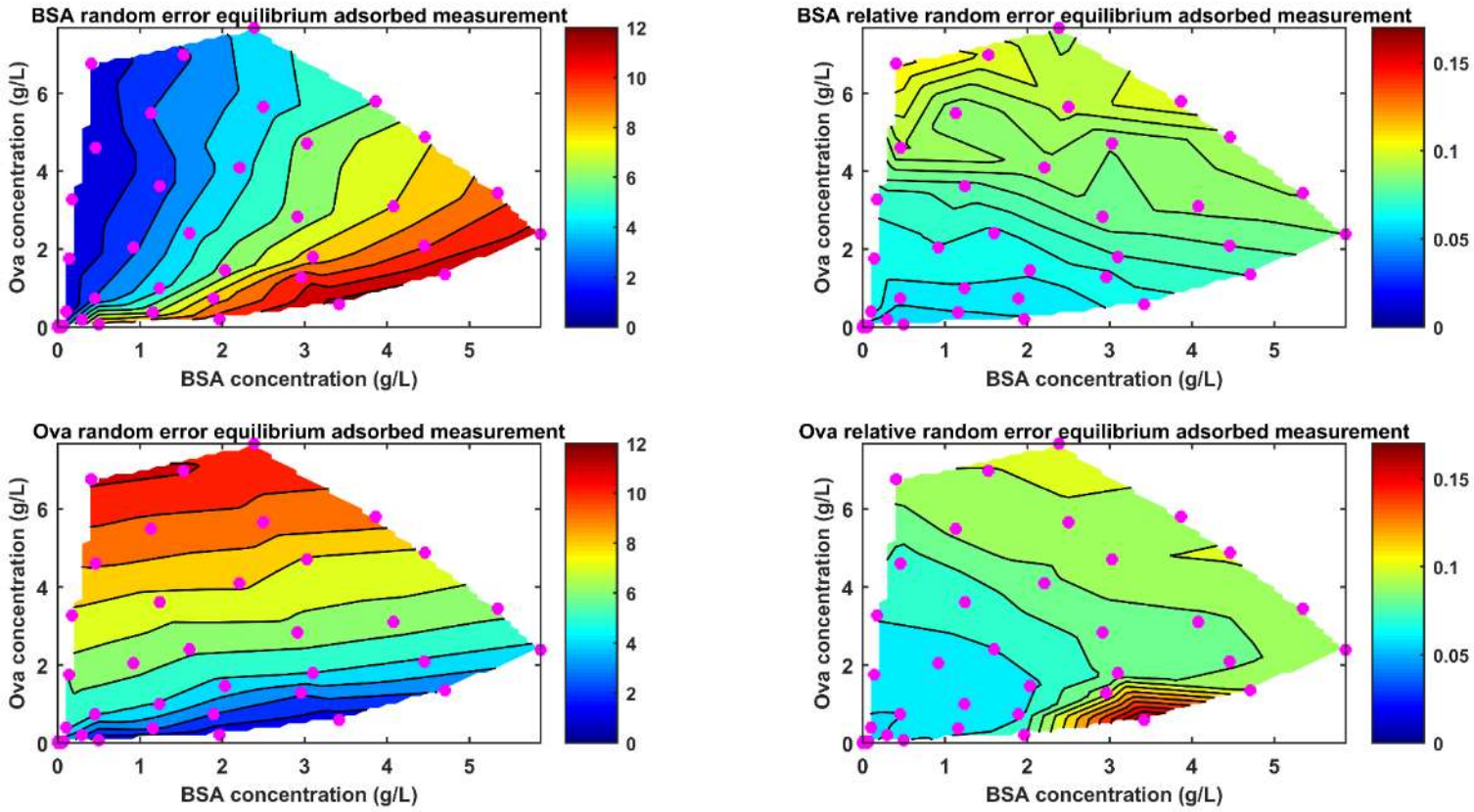

Figure 6-9 Contour plots showing propagated random error in adsorbed concentration calculated using q for BSA-Ova binary isotherms at pH 9 0 mM NaCl. Top 2 panels show propagated error for BSA and bottom 2 for Ova. Left hand panels show propagated error in adsorbed concentration mg/mL resin and right hand panels show relative error in adsorbed concentration.

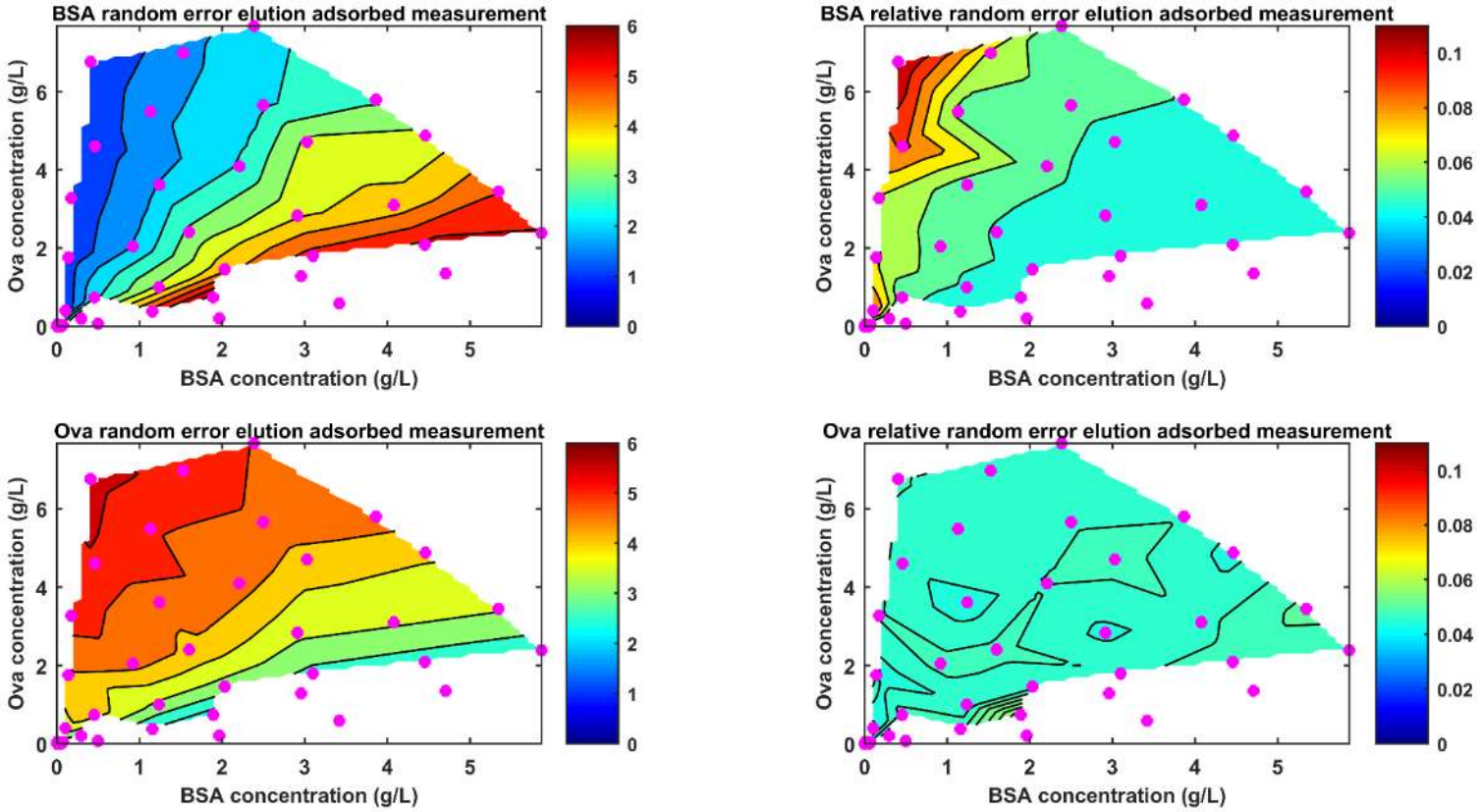

Figure 6-10 Contour plots showing propagated random error in adsorbed concentration calculated using q for BSA-Ova binary isotherms at pH 9 0 mM NaCl. Top 2 panels show propagated error for BSA and bottom 2 for Ova. Left hand panels show propagated error in adsorbed concentration mg/mL resin and right hand panels show relative error in adsorbed concentration.*

6.1.2.3 50 mM NaCl isotherm

The binary isotherm is presented in Figure 6-11 along with agreement between the two measured of adsorbed concentration q and q^*. The BSA and Ova isotherms in Figure 6-11A and B in the presence of little competition retain a similar shape to what was observed for the single component isotherms in the same buffer conditions. BSA keeps a highly rectangular appearance with very little increase in liquid concentration with increasing adsorbed concentration until the plateau is reached. Ova seems to have a less rectangular shape with increases in liquid concentration with increasing adsorbed concentration before the plateau is reached in the non-competitive region of the isotherm.

The isotherms show significant competition and displacement for both components. BSA which is the most retained species is effectively displaced by small quantities of Ova with 3 mg/mL starting concentration of Ova deceasing the plateau and decreasing the rectangularity of the isotherm. This similar to what was observed at 0 mM NaCl except the displacement of Ova has been decreased marginally at 50 mM NaCl. Ova has a lower saturation capacity than BSA and is the less retained species. Moreover, BSA has become more effective at displacing than at 0 mM NaCl. This ties into the column experiments and SCIs shown for BSA and Ova, Ova is the lesser retained species and elutes before BSA in column runs, it also shows increased sensitivity to NaCl in SCIs, so as the NaCl level increases BSA becomes more effective at displacing Ova and Ova becomes less effective at displacing BSA. Again there seems to be little additional effect of displacement above 9 mg/mL starting concentration for either protein which was also true at 0 mM NaCl. Figure 6-11C and D show generally good agreement between adsorbed concentrations calculated from elution and via mass balance. As there is good agreement between q and q^* and error bars are generally under good control systematic and random error are presented in Appendix A: Additional binary isotherm datasets.

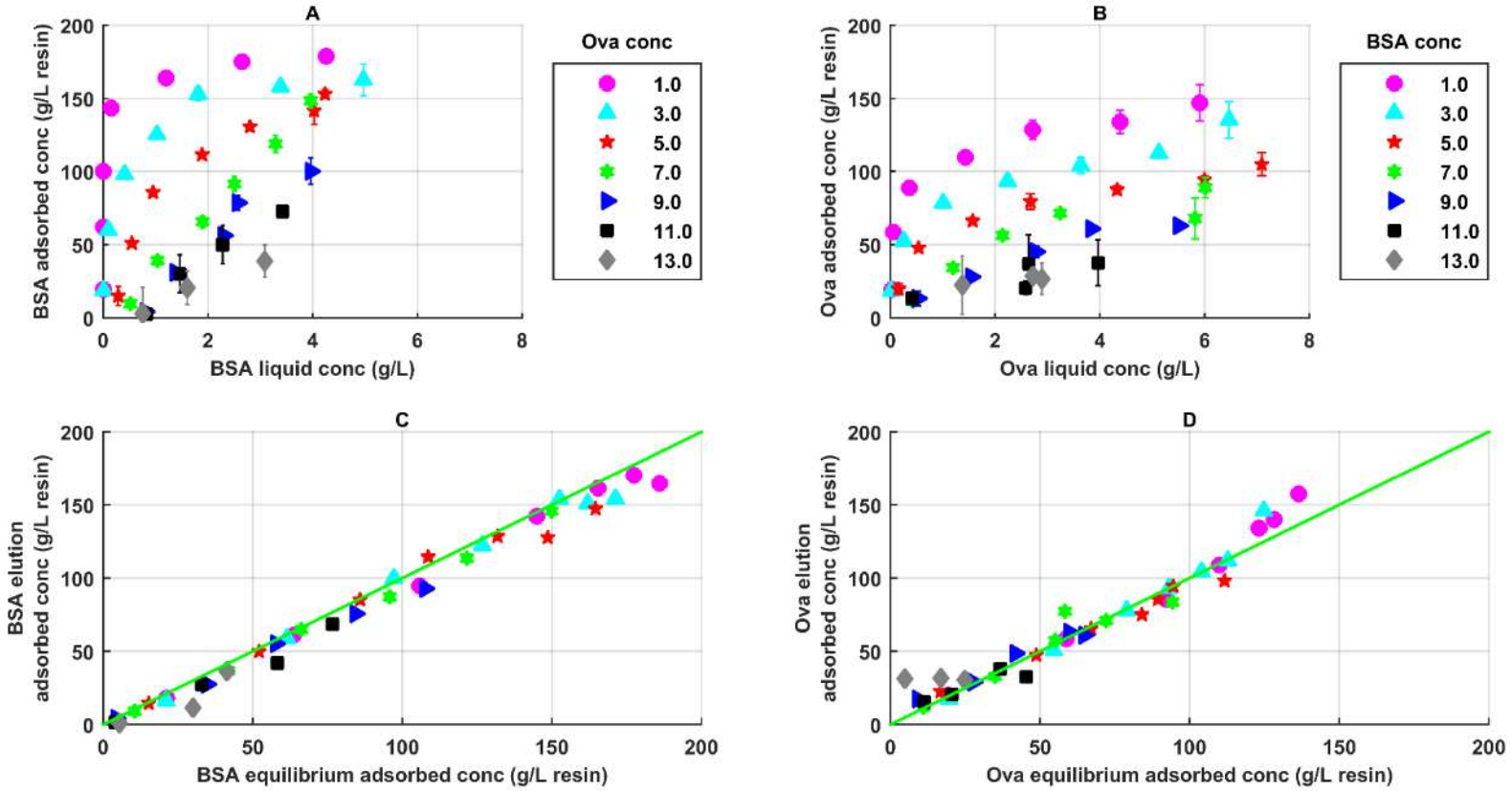

Figure 6-11 BSA-Ova binary isotherm pH 9 50 mM NaCl. Panels A & C present BSA liquid and adsorbed concentrations, B & D represent Ova data. Panels A and B present the binary isotherms at different starting concentrations of the competitor displayed in the figure legend in mg/mL, the liquid concentration is the average of q and q after outlier rejection. Error bars in panels A and B represent experimental standard deviation. Panels C and D present agreement between q and q*with a line of parity in green.*

6.1.2.4 100 mM NaCl isotherm

The binary isotherm is presented in Figure 6-12 along with agreement between q and q^*. The shape of the BSA and Ova isotherms in Figure 6-12A and B in the presence of little competition retain a similar shape to what was observed for the single component isotherms in the same buffer conditions. At this increased NaCl level BSA begins to lose its highly rectangular appearance. The rectangularity of the Ova isotherm is decreasing further with increasing NaCl.

With increasing NaCl the competitive nature of the isotherms is changing. Ova is not particularly effective at displacing BSA with significant suppression only happening at 5-7 mg/mL starting concentration of Ova. Ova is now very effectively displaced by BSA with significant suppression at only 3 mg/mL starting concentration of BSA. Again there seems to be little additional effect of displacement above 9 mg/mL starting concentration for either protein which was also true at 0 and 50 mM NaCl. Figure 6-12C shows good agreement between adsorbed concentrations calculated from elution and via mass balance for BSA but Ova seems to show some disagreement. There is a tendency for the eluted concentration to underestimate the adsorbed concentration in comparison with the equilibrium concentration, propagation of systematic errors will explore this disagreement in the next section.

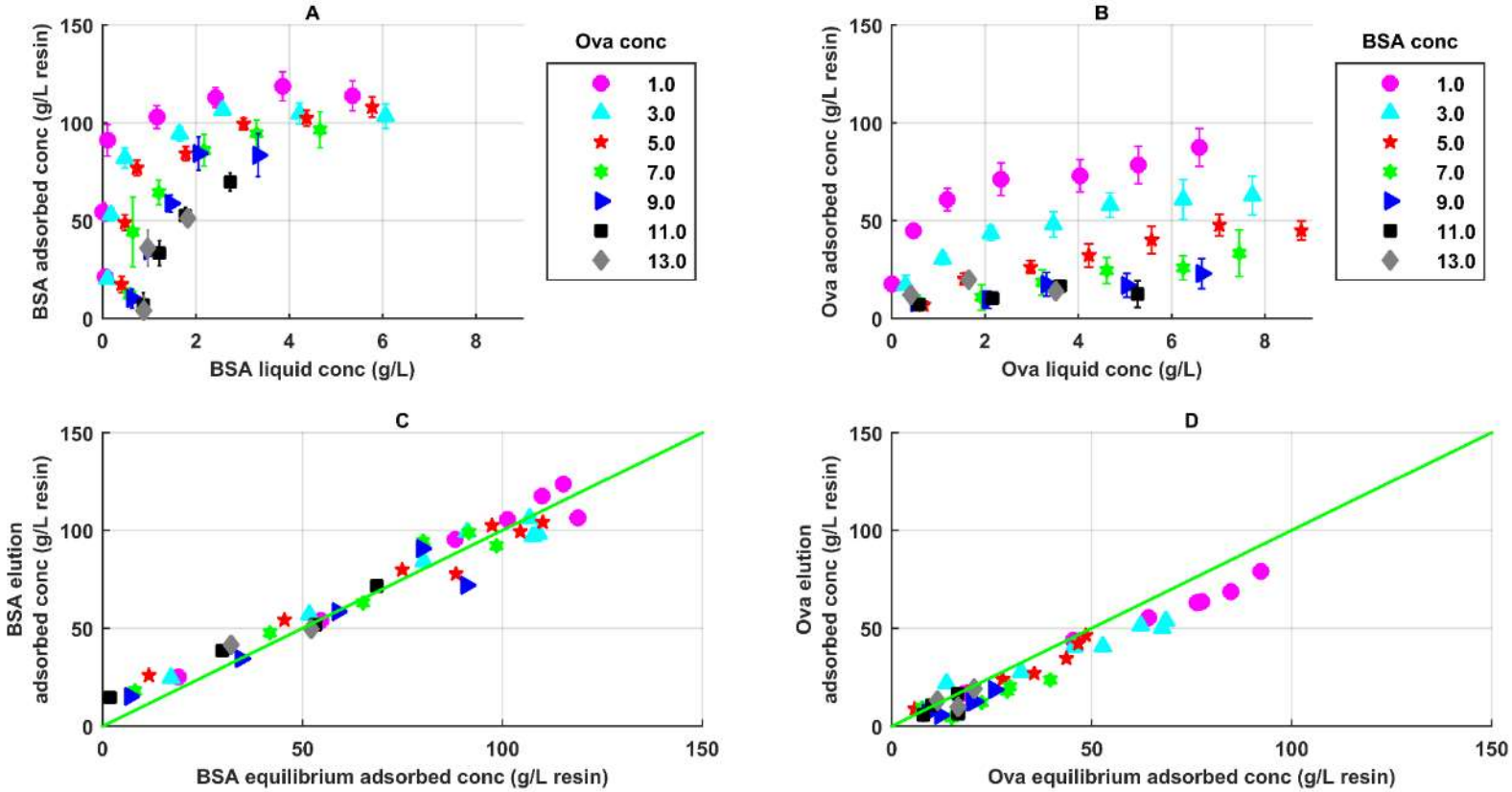

Figure 6-12 BSA-Ova binary isotherm pH 9 100 mM NaCl. Panels A & C present BSA liquid and adsorbed concentrations, B & D represent Ova data. Panels A and B present the binary isotherms at different starting concentrations of the competitor displayed in the figure legend in mg/mL, the liquid concentration is the average of q and q after outlier rejection. Error bars in panels A and B represent experimental standard deviation. Panels C and D present agreement between q and q*with a line of parity in green.*

6.1.2.5 100 mM NaCl error propagation

Figure 6-13 shows the BSA-Ova binary isotherm at pH 9 100 mM NaCl as a contour plot, this has been included as a reference to facilitate comparison to propagated error contour plots.

*6.1.2.5.1 Systematic error in q and q**

Figure 6-14 shows the propagated systematic error in *q,* the left hand plots show that the propagated error reaches a maximum of ≈ 3 mg/mL resin for BSA and ≈ 5 mg/mL for Ova and that there is no pattern within components, there is however a pattern across components which was discussed previously. Relative error as calculated in equation 6-5 shows BSA reaches a maximum of approximately 0.4 or 40% and Ova of 0.2 or 20%. Although 40% relative error for BSA is high it is localised to regions of very low adsorption where the denominator is controlling in equation 6-5. Ova has some regions of high relative systematic error where the error in mg/mL resin is also high meaning the high relative error is driven by high error in mg/mL rather than just a low adsorbed concentration.

Figure 6-15 shows the propagated systematic error in *q** for BSA-Ova binary isotherm at pH 9 100 mM NaCl. Left hand plots show that the propagated error reaches a maximum of 2 mg/mL resin for BSA and 3 mg/mL for Ova and that there is no pattern. Relative error shows that BSA reaches a maximum of 0.06 or 6% and Ova 0.25 or 25%. The BSA maximum relative error is highly localised to a region of low adsorption where adsorbed concentration is low. The region of high relative systematic error for Ova is less localised and a region of low adsorption coincides with a region of higher systematic error in mg/mL resin.

*6.1.2.5.2 Random error in q and q**

Figure 6-16 shows the propagated random error in *q*, the left hand plots shows that the propagated error reaches a maximum of 10 mg/mL resin for BSA and Ova. There is a strong pattern of increasing random error with increasing assayed concentration for the protein being presented in the left hand plots. This pattern was explained in *6.1.2.2 0 mM error propagation*. In terms of relative error calculated in equation 6-5

BSA reaches a highly localised maximum of 1.2 or 120% at a region of low adsorption where the denominator is controlling, the error in terms if mg/mL of resin is low at around 1 mg/mL resin in this region. For Ova the relative random error reaches a maximum of 0.3 or 30% and is more generally spread across the plot although it is still in regions of low Ova adsorption and the error in terms of mg/mL of resin is still quite low around 2-6 mg/mL resin.

Figure 6-17 shows the propagated random error in q^*, the left hand plots show that the propagated error reaches a maximum of 7 and 5 mg/mL resin for BSA and Ova respectively. The random error is strongly linked to the adsorbed concentration, as previously discussed. Relative error calculated in equation 6-5 shows BSA reaches a highly localised maximum of 0.1 or 10%. For Ova the relative random error reaches a maximum of 0.23 or 23%, the region of high relative random error for Ova is in an area of low error in terms of mg/mL resin but is caused by the low adsorbed concentration as the denominator is controlling.

6.1.2.5.3 Summary of propagated error

Analysis of these propagated errors suggests why measuring multicomponent isotherms is challenging in the presence of a modifier, such as NaCl which weakens the strength of interaction. The errors in terms of mg/mL of resin have not increased dramatically from 0 NaCl to 100 mM NaCl but the relative error are increasing particularly for Ova which is the less retained species as the adsorbed concentration drops. There is some disagreement between the equilibrium and elution measures of adsorbed concentration for Ova at this NaCl concentration. Whilst the reason for this may partially be explained by the systematic error propagation of the isotherms as regions of low adsorbed concentration have higher relative systematic error there is also disagreement is regions of high adsorbed concentration when looking at Figure 6-12C. This disagreement is not totally accounted for by propagation of systematic errors as according to the error propagation the relative systematic error in these regions should be low at both equilibrium and elution measurements. As this disagreement is not totally explained both estimates of q and q^* will remained

averaged across one another as it is unclear which one is associated with least systematic error.

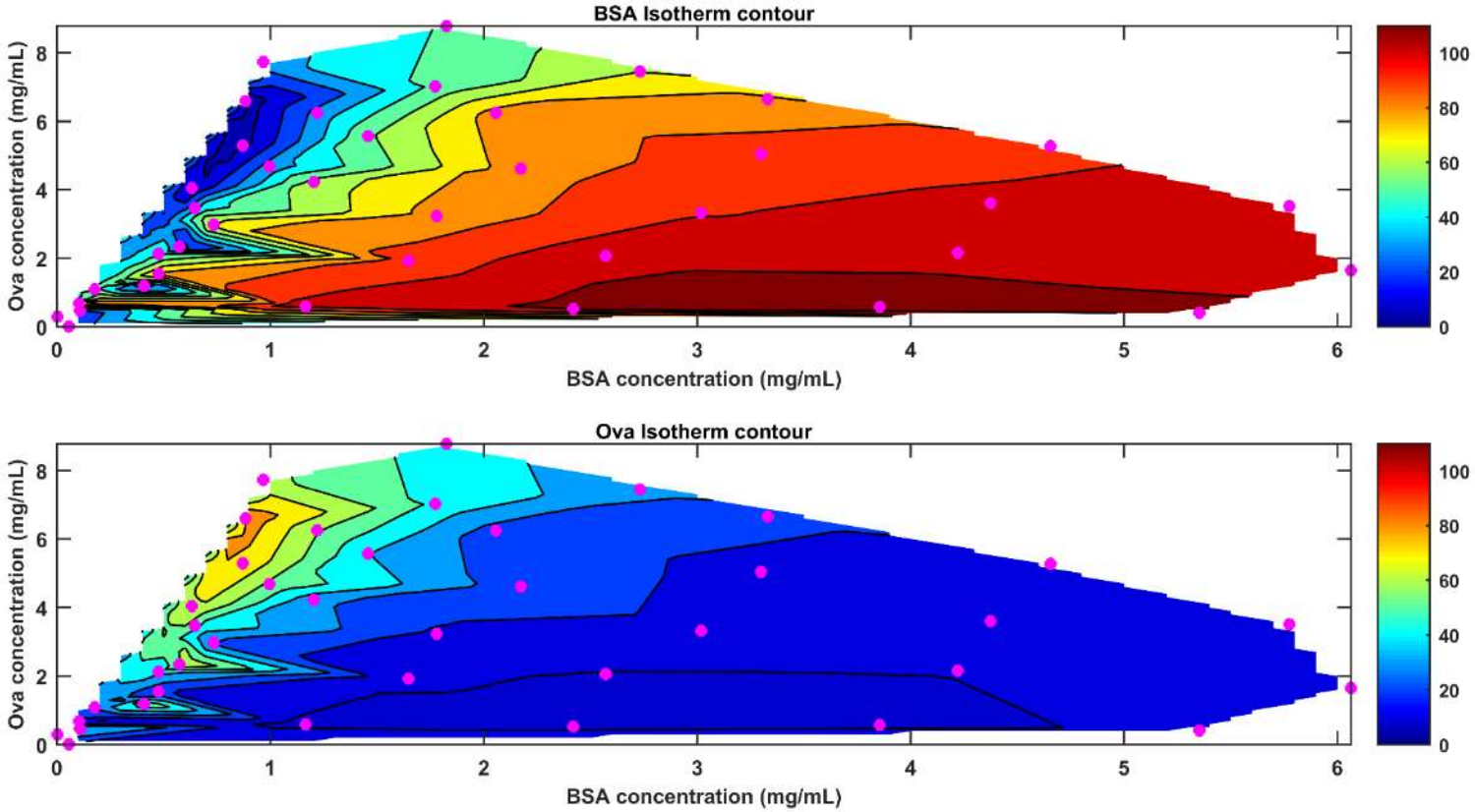

Figure 6-13 Binary isotherm for BSA-Ova pH 9 100 mM NaCl. Isotherm has been drawn as a contour plot as a reference for error propagation plots. x and y axis display average equilibrium concentrations of BSA and Ova measured directly after outlier rejection, contours display adsorbed concentration of BSA in the top panel and Ova in the bottom panel, these have been taken as averages of q and q.*

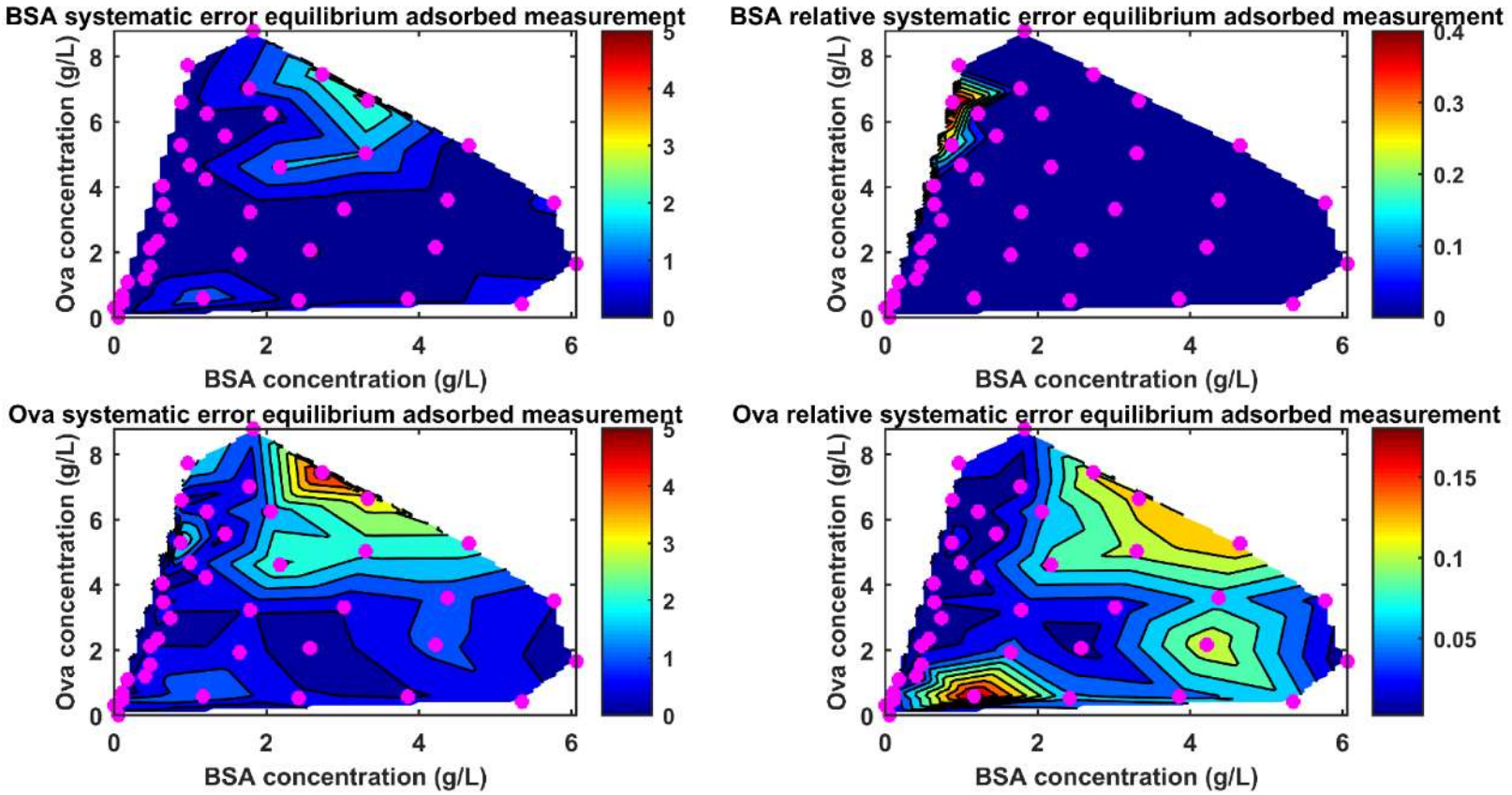

Figure 6-14 Contour plots showing propagated systematic absolute error in adsorbed concentration calculated using q for BSA-Ova binary isotherms at pH 9 100 mM NaCl. Top 2 panels show propagated error for BSA and bottom 2 for Ova. Left hand panels show propagated error in adsorbed concentration mg/mL resin and right hand panels show relative error in adsorbed concentration.

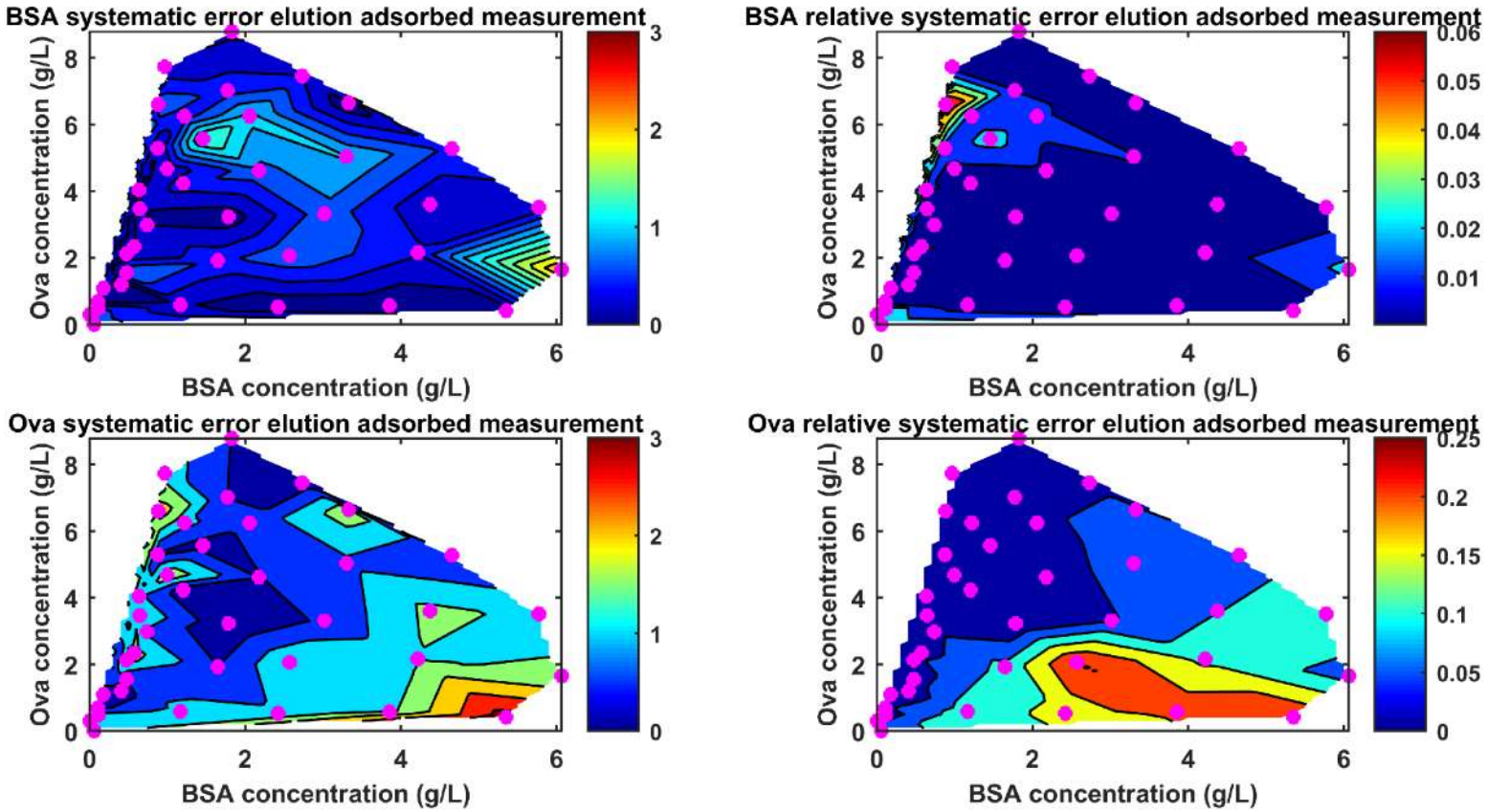

*Figure 6-15 Contour plots showing propagated systematic absolute error in adsorbed concentration calculated using q*for BSA-Ova binary isotherms at pH 9 100 mM NaCl. Top 2 panels show propagated error for BSA and bottom 2 for Ova. Left hand panels show propagated error in adsorbed concentration mg/mL resin and right hand panels show relative error in adsorbed concentration.*

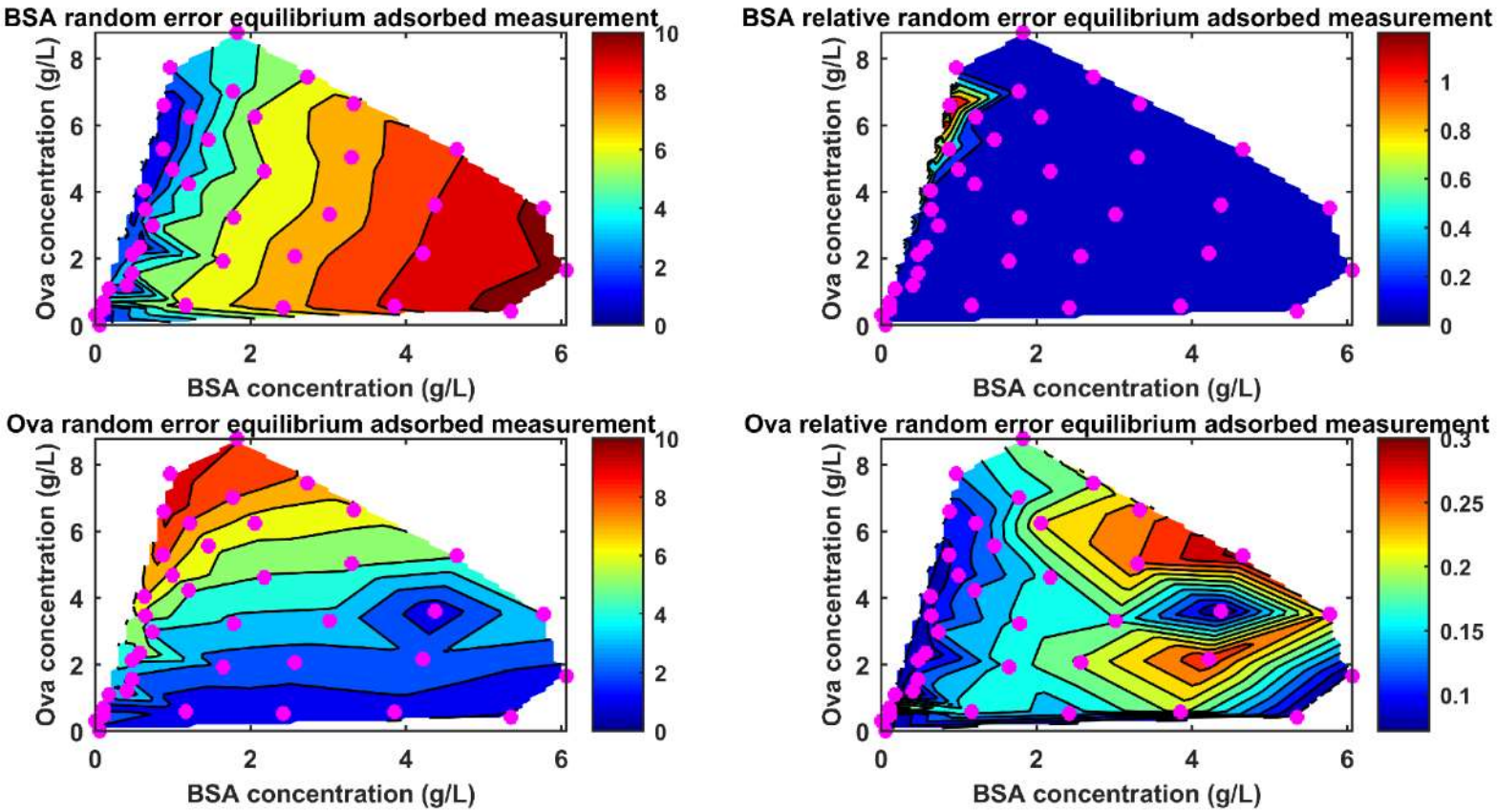

Figure 6-16 Contour plots showing propagated random error in adsorbed concentration calculated using q for BSA-Ova binary isotherms at pH 9 100 mM NaCl. Top 2 panels show propagated error for BSA and bottom 2 for Ova. Left hand panels show propagated error in adsorbed concentration mg/mL resin and right hand panels show relative error in adsorbed concentration.

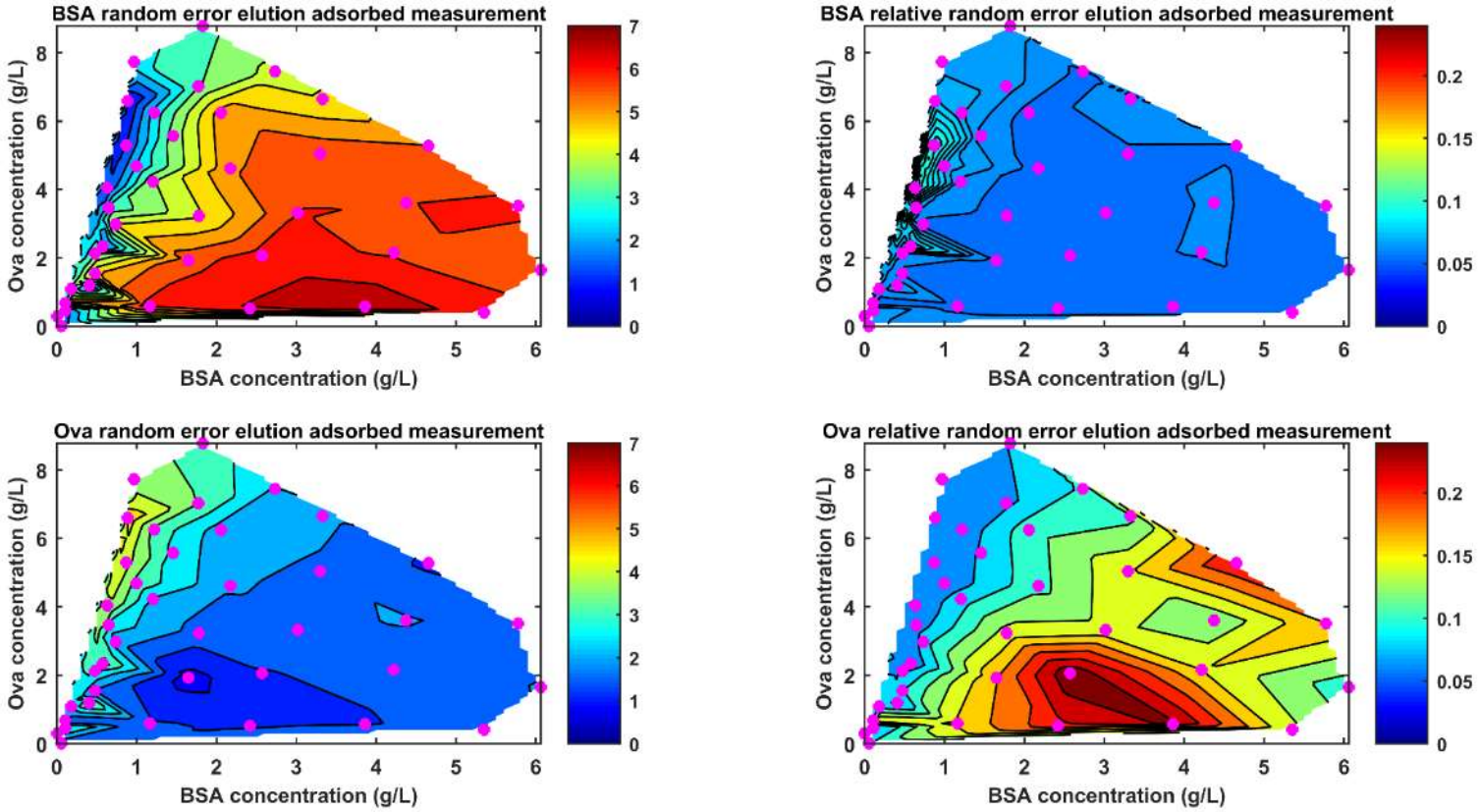

Figure 6-17 Contour plots showing propagated random error in adsorbed concentration calculated using q for BSA-Ova binary isotherms at pH 9 100 mM NaCl. Top 2 panels show propagated error for BSA and bottom 2 for Ova. Left hand panels show propagated error in adsorbed concentration mg/mL resin and right hand panels show relative error in adsorbed concentration.*

6.1.2.6 200 mM NaCl isotherm

The binary isotherm is presented in Figure 6-18 along with agreement between q and q^*. Both BSA and Ova have been significantly suppressed by the NaCl with rectangular behaviour being entirely lost at the lowest starting concentrations of competitor. There seems to be no additional suppression of BSA with increasing Ova concentration at this high NaCl level. The Ova isotherm is rather noisy with significant error bars and isotherms with different starting competitor concentrations crossing over, this will be discussed further in the error propagation section below.

Figure 6-18C and D show good agreement between adsorbed concentrations calculated from elution and via mass. There is a wide spread of points around the line of parity in plot Figure 6-18D; this will be discussed in more detail in the error propagation section.

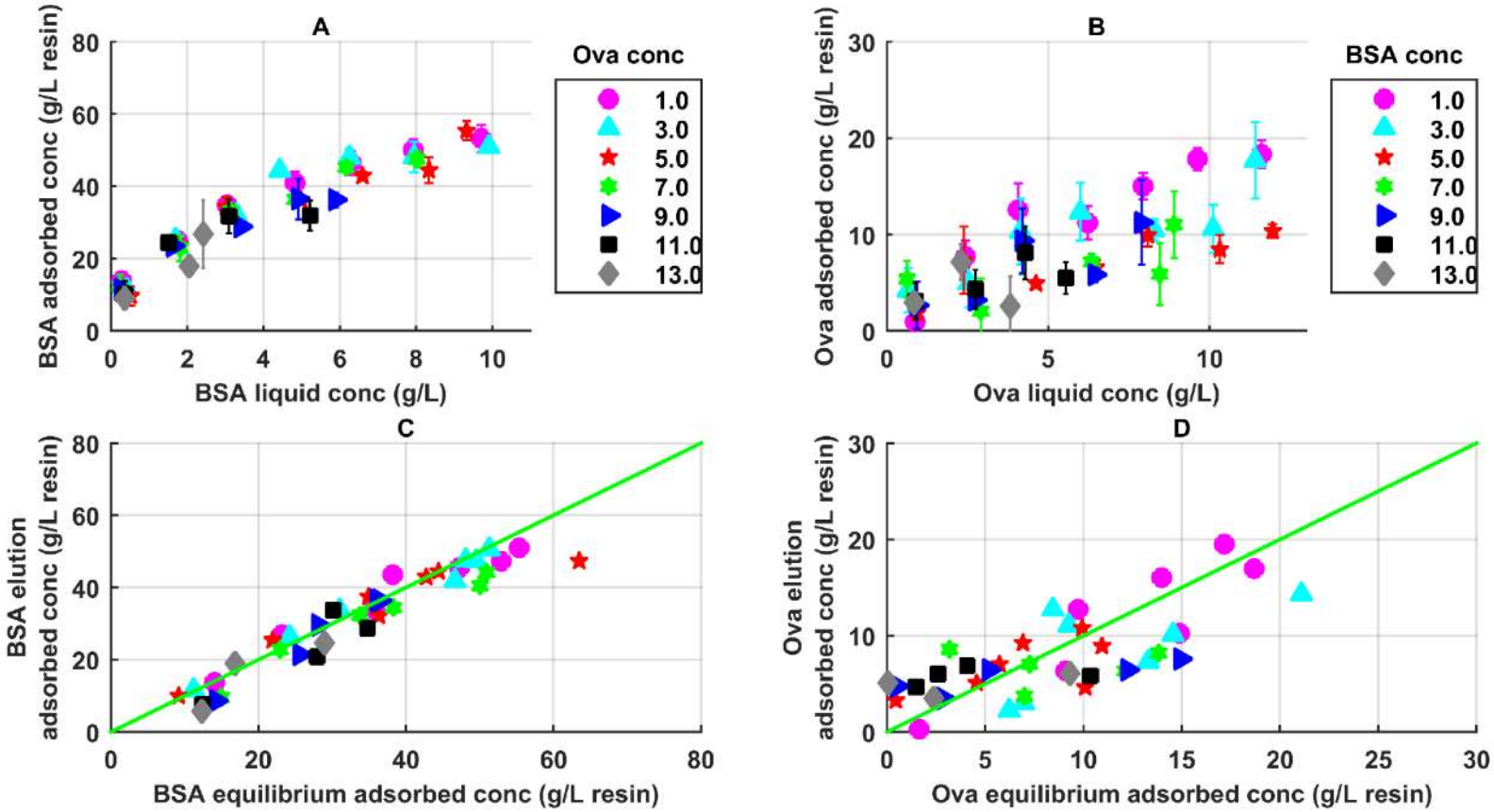

Figure 6-18 BSA-Ova binary isotherm pH 9 200 mM NaCl. Panels A & C present BSA liquid and adsorbed concentrations, B & D represent Ova data. Panels A and B present the binary isotherms at different starting concentrations of the competitor displayed in the figure legend in mg/mL, the liquid concentration is the average of q and q after outlier rejection. Error bars in panels A and B represent experimental standard deviation. Panels C and D present agreement between q and q*with a line of parity in green.*

6.1.2.7 200 mM NaCl error propagation

Figure 6-19 shows the BSA-Ova binary isotherm at pH 9 200 mM NaCl as a contour plot. This has been included as a reference to facilitate comparison to propagated error contour plots. The BSA isotherm contour plot shows the expected trends with adsorbed concentration of BSA increasing with increasing liquid concentration of BSA. There is minimal displacement of BSA by Ova is also shown in Figure 6-18. The Ova contour isotherm shows much less clear trends, just as in Figure 6-18.

*6.1.2.7.1 Systematic error in q and q**

Figure 6-20 shows the propagated systematic error in q, the left hand plots show that the propagated error reaches a maximum of 3 mg/mL resin for BSA and 7 mg/mL for Ova and that there is no pattern within the components, there is however a pattern across the components which was discussed previously. BSA relative errors reach a maximum of approximately 0.1 or 10% and Ova of 2 or 200%. There is a missing corner of the contour plot in Figure 6-20 relative systematic error of Ova, which was intentionally omitted as it showed vert high relative error, which made the rest of the contour plot difficult to display. 10% systematic error for BSA is reasonable especially considering it is caused by systematic error of 3 mg/mL of resin. The relative systematic error for Ova is quite high across the plot as adsorbed concentration is so low. Systematic error in mg/mL remains reasonable at a maximum of 7 mg/mL resin.

Figure 6-21 shows the propagated systematic in q^* for the BSA-Ova binary isotherm at pH 9 200 mM NaCl. Left hand plots show that the propagated error reaches a maximum of 1 mg/mL resin for BSA and 3 mg/mL for Ova. BSA reaches a maximum of 0.07 or 7% and Ova 0.09 or 9% relative error. The BSA maximum relative error is contained to a region of low adsorption and the region of high relative systematic error for Ova is localised to one corner of low adsorption.

*6.1.2.7.2 Random error in q and q**

Figure 6-22 shows the propagated random error in q, the left hand plots show that the propagated error reaches a maximum of 10 mg/mL resin for BSA and Ova. There is a strong pattern of increasing random error with increasing assayed concentration

for the protein being presented in the left hand plots. BSA relative error reaches a highly localised maximum of 0.20 or 20%, this is driven by the greater random error in mg/mL resin as the adsorbed concentration is relatively high here. For Ova the relative random error reaches a maximum of 3 or 300%. Hot spots of high relative random error are spread across the plot, a corner of the plot has been intentionally omitted as very high values in this corner were affecting the remainder of the contour plot. The error is only 3 mg/mL but the high values of relative error are driven by the low adsorbed concentration for this isotherm.

Figure 6-23 shows the propagated random error in *q** for BSA-Ova binary isotherm at pH 9 200 mM NaCl. Left hand plots show that the propagated error reaches a maximum of 3 and 2 mg/mL resin for BSA and Ova respectively. The random error is strongly linked to the adsorbed concentration, the reasons for which were described in *6.1.2.2 0 mM error propagation.* BSA reaches a highly localised maximum of 0.1 or 10% relative error. For Ova the relative random error reaches a maximum of 0.3 or 30%, the region of high relative random error for Ova is in an area of low error in terms of mg/mL resin but is caused by the low adsorbed concentration as the denominator is controlling. The rest of the isotherm shows random error of around 15% or less. This is in contrast to values of around 50% for the measurement at equilibrium.

6.1.2.7.3 Summary of propagated error

The analysis of propagated error shows that random error of Ova is much greater for *q* than *q**. As discussed at 100 mM NaCl the high levels of relative error are driven by the low adsorbed concentrations experienced at these high NaCl levels demonstrating why it's difficult to study more weakly interacting isotherms. The high levels of relative random error for Ova when measured via equilibrium measurement may explain why there is broad spread around the parity plot in Figure 6-18. Because of this high level of random error of *q* for Ova at 200 mM in the BSA-Ova binary the isotherm this will be presented again but only showing the *q** data which had much lower levels of propagated random error displayed.

6.1.2.8 200 mM NaCl Ova half of isotherm with *q** data only

The Ova part of the isotherm is presented with only *q** data in Figure 6-24. Figure 6-24A does shows error bars of reduced size compared to Figure 6-18B, there is still some crossing at the low liquid concentration levels of Ova, although it is in the range of mg/mL resin which is low. The isotherm contour in Figure 6-24B shows much clearer trends than Figure 6-19B.

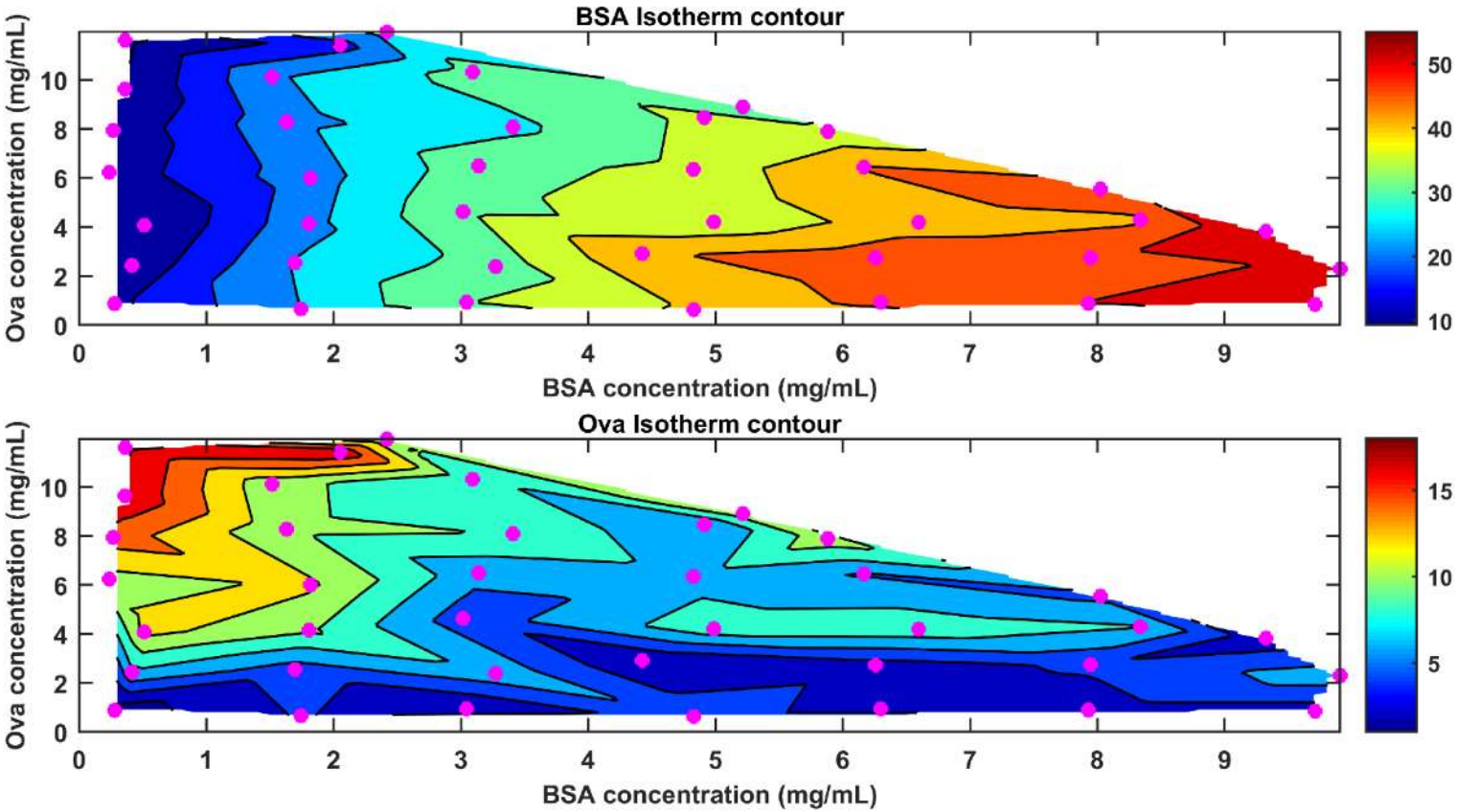

Figure 6-19 Binary isotherm for BSA-Ova pH 9 200 mM NaCl. Isotherm has been drawn as a contour plot as a reference for error propagation plots. x and y axis display average equilibrium concentrations of BSA and Ova measured directly after outlier rejection, contours display adsorbed concentration of BSA in the top panel and Ova in the bottom panel, these have been taken as averages of q and q.*

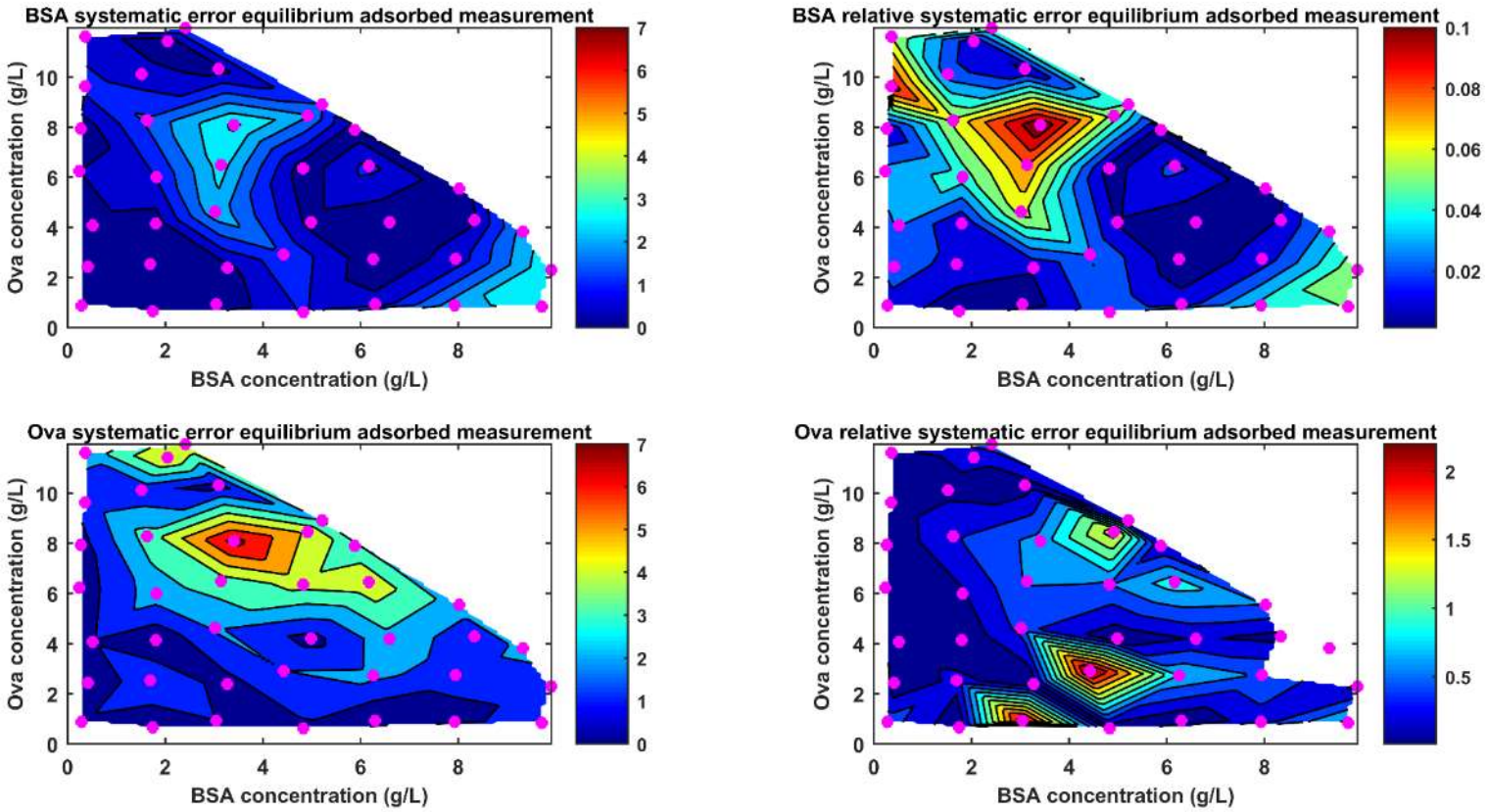

Figure 6-20 Contour plots showing propagated systematic absolute error in adsorbed concentration calculated using q for BSA-Ova binary isotherms at pH 9 200 mM NaCl. Top 2 panels show propagated error for BSA and bottom 2 for Ova. Left hand panels show propagated error in adsorbed concentration mg/mL resin and right hand panels show relative error in adsorbed concentration.

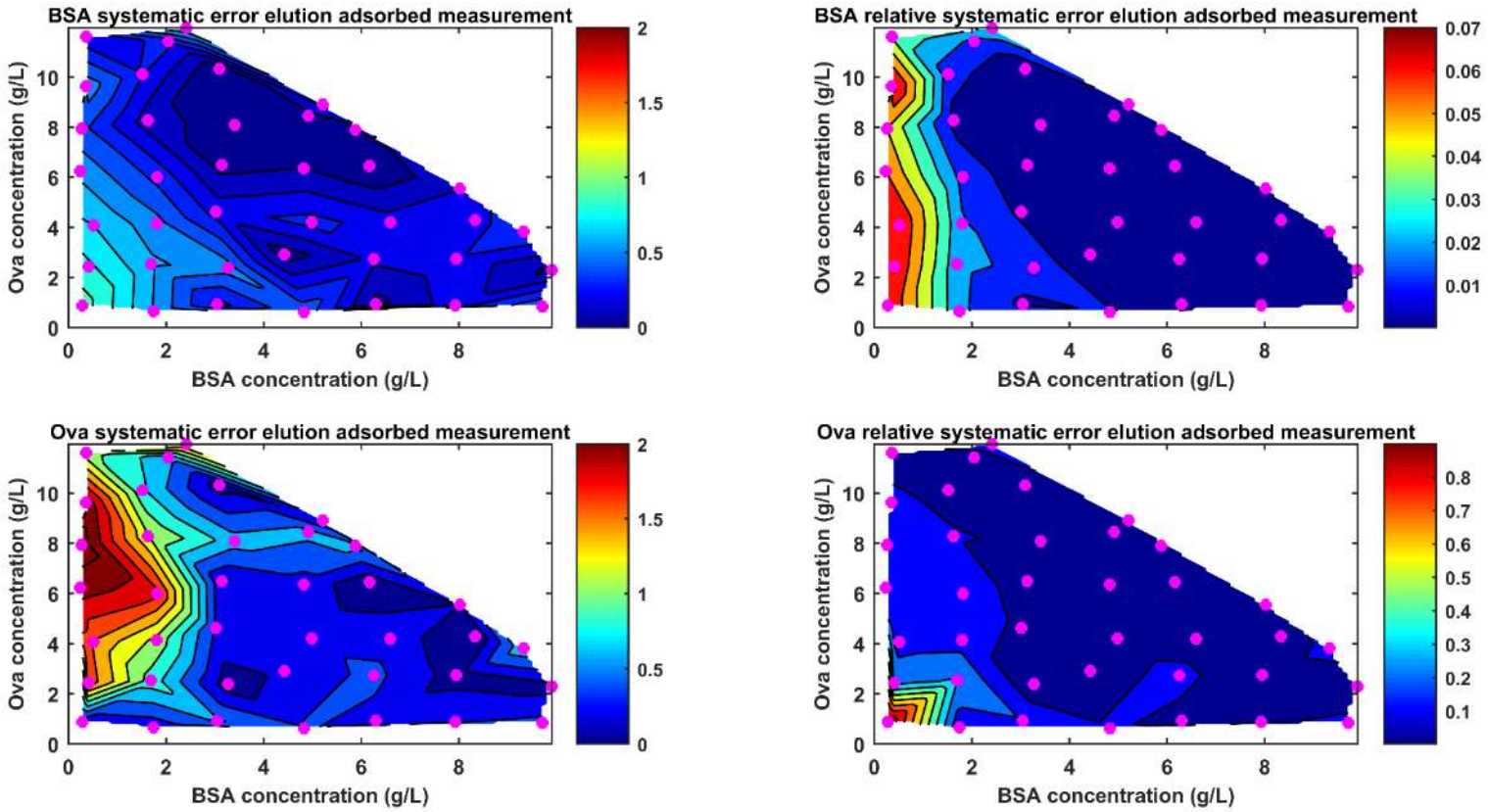

*Figure 6-21 Contour plots showing propagated systematic absolute error in adsorbed concentration calculated using q*for BSA-Ova binary isotherms at pH 9 200 mM NaCl. Top 2 panels show propagated error for BSA and bottom 2 for Ova. Left hand panels show propagated error in adsorbed concentration mg/mL resin and right hand panels show relative error in adsorbed concentration.*

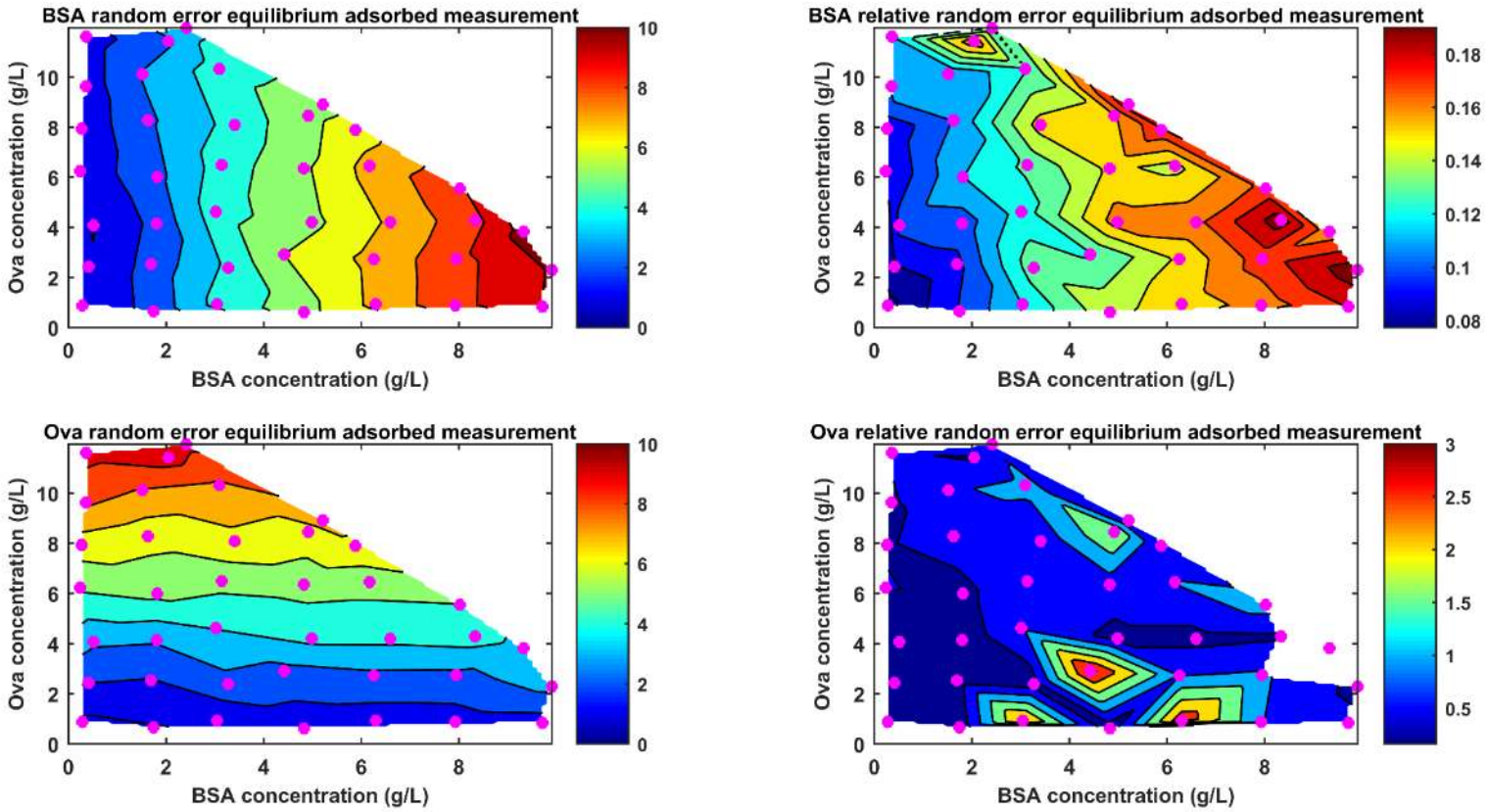

Figure 6-22 Contour plots showing propagated random error in adsorbed concentration calculated using q for BSA-Ova binary isotherms at pH 9 200 mM NaCl. Top 2 panels show propagated error for BSA and bottom 2 for Ova. Left hand panels show propagated error in adsorbed concentration mg/mL resin and right hand panels show relative error in adsorbed concentration.

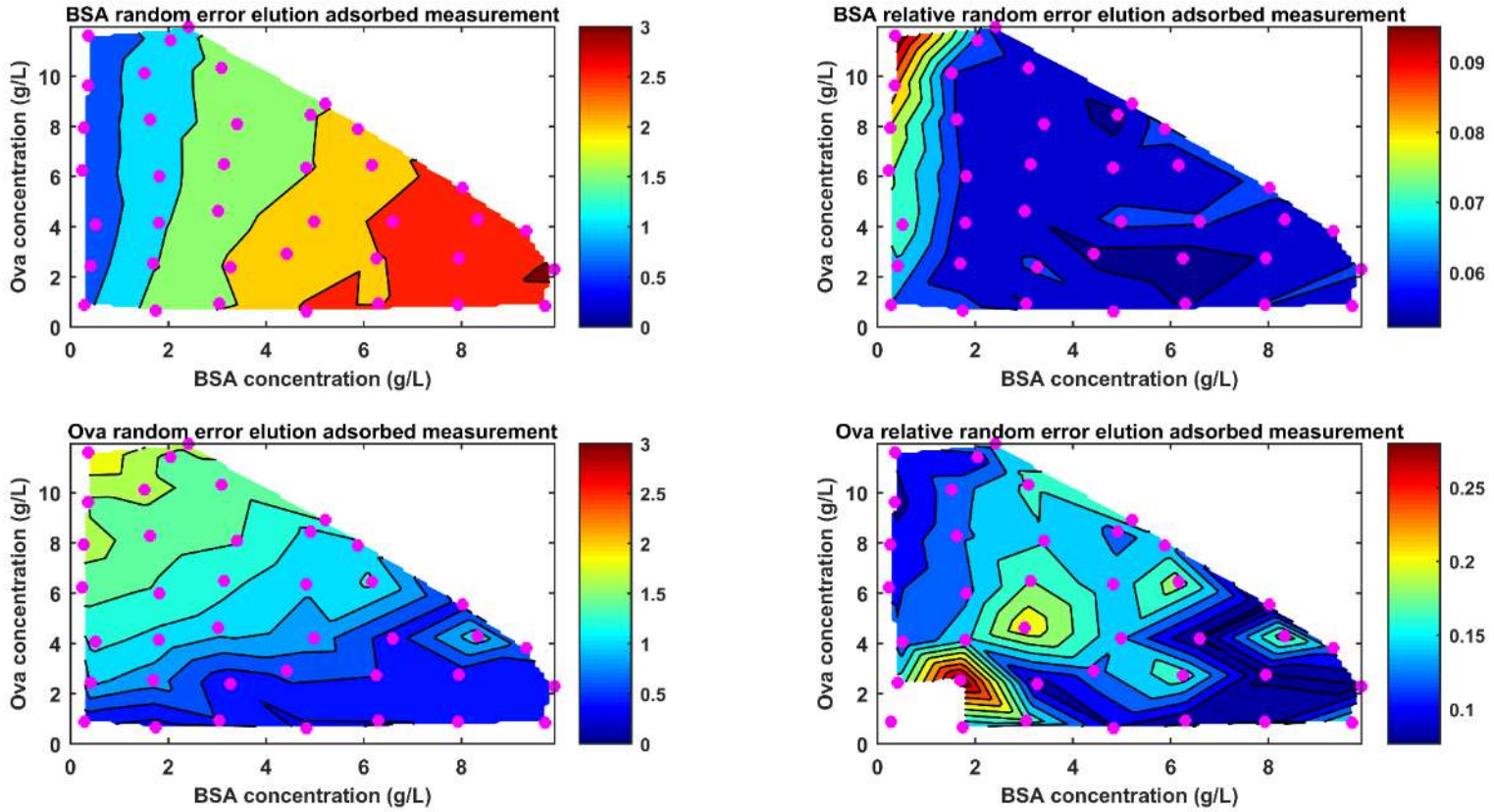

Figure 6-23 Contour plots showing propagated random error in adsorbed concentration calculated using q for BSA-Ova binary isotherms at pH 9 50 mM NaCl. Top 2 panels show propagated error for BSA and bottom 2 for Ova. Left hand panels show propagated error in adsorbed concentration mg/mL resin and right hand panels show relative error in adsorbed concentration.*

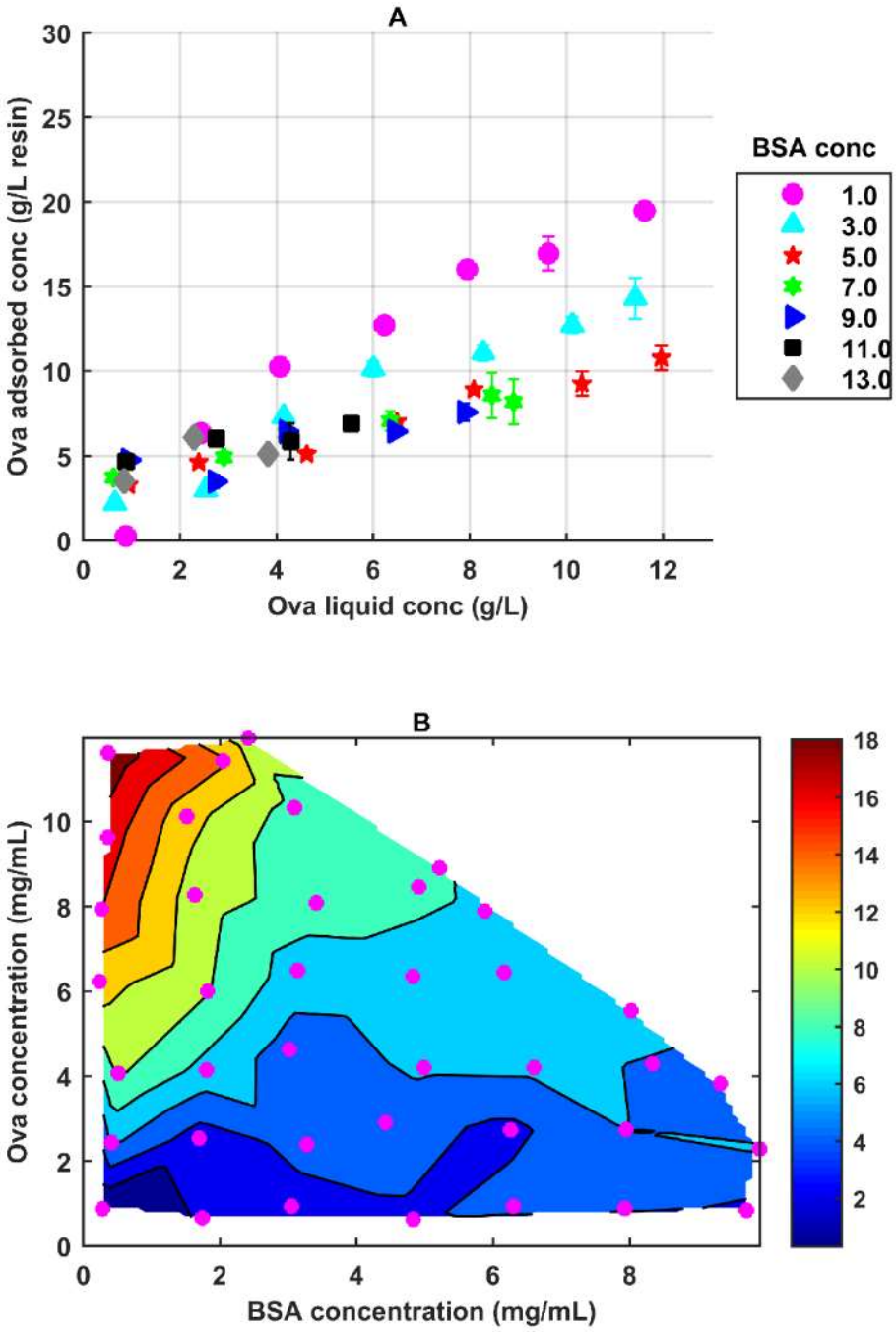

Figure 6-24 Ova half of BSA-Ova binary isotherms at 200 mM NaCl using q only for adsorbed concentration estimate. A shows the isotherm plotted as if a series of single component isotherm wither different levels of start concentration of BSA represented in the legend in mg/mL. B shows the isotherm as a contour plot.*

6.1.3 Summary of BSA-Ova binary isotherm pH 9 and the propagation of errors in *q* and *q**

At low concentrations of NaCl (0 and 50 mM) Ova and BSA show significant mutual displacement with Ova displacing BSA more effectively than vice versa. Additionally there seems to be little additional displacement above 9 mg/mL for the starting concentration of the competitor. As the NaCl level increases the isotherm shows similar behaviours as to what was observed in the SCI with rectangular behaviour decreasing and the plateau of the isotherm lowering. BSA becomes increasing effective at displacing Ova. Conversely Ova becomes decreasingly effective at displacing BSA with increasing NaCl.

There is no clear pattern of systematic error for either BSA or Ova isotherms when comparing systematic error within the component. However, when considering systematic error across components with regard to *q* there is tendency for errors to be higher in the same regions for both components. As discussed previously, the UV spectra-PLS method has a tendency to compensate during estimation meaning when one compound is overestimated the other is underestimated. As the plots show absolute error without information as to the direction of the error it presents error as high for both compounds at the same location. These patterns are not so apparent when observing *q** as there are multiple fractions being analysed and contributing to the systematic error which confounds any relationship.

There are clear patterns in random error for both *q* and *q**: The pattern in random error for *q* is driven by the random error in assayed equilibrium concentration which is modelled linearly on the assayed concentration as shown in Figure 6-2. Additionally the C_0 random error increases proportionally with C_0.

When calculating adsorbed concentration via elution the random error increases with adsorbed concentration. This is for the same reason as discussed in the above paragraph. As the adsorbed concentration increases the assayed concentration in elution fraction 1 increases which in turn is related with an increased random error. The different way in which random error presents when calculating

adsorbed concentration from equilibrium vs elution can be exploited to reduce random error in specific circumstances.

There are similar levels of systematic error within the protein and across the NaCl levels in terms of error in mg/mL resin. There are also similar levels of random error in the q measurement across NaCl levels. However, as the random error in q^* in mg/mL is linked to the adsorbed concentration it decreases as NaCl level increases and adsorbed concentration decreases. This presents an advantage when studying isotherms with small adsorbed concentrations at high NaCl levels as the relative random error is high as the denominator becomes controlling in equation 6-5 as in the case of Ova at 200 mM NaCl.

Isotherms with small adsorbed concentration are generally more challenging to study because the relative errors become large when the adsorbed concentration being small in equation 6-5. Elution can offer an attractive alternative method of quantification when conditions become challenging.

6.2 Discussion and summary

There is only a limited amount of binary protein isotherms in the literature. The examples which do exist in the literature either focus on the same cytochrome-c lysozyme binary system, which is easy to study due to the presence of a chromophore in cytochrome-c (Cyc-C) (Baumann et al., 2016; Xu & Lenhoff, 2009). Other datasets have been somewhat error prone with noise or implausible behaviour apparent in the data (Liang, Fieg, & Jakobtorweihen, 2015; Yang, 2016; Q. Zhang, Schimpf, Lu, Lin, & Yao, 2016).

An extensive literature search suggests a detailed study of error propagation in isotherm measurement has not been published, which is why the propagation of systematic and random errors has been presented for a binary isotherms at a range of NaCl levels. Additional binary isotherm datasets and their error propagation are presented in Appendix A: Additional binary isotherm datasets. These data sets display different competitive behaviour to the BSA-Ova dataset discussed here. In the BSA-Con dataset Con is very effectively displaced by BSA but Con has little effect at displacing BSA. As the NaCl concentration increases Con becomes decreasingly effective at displacing BSA. Similar behaviour is observed for the Ova-Con dataset with Con ineffective at displacing Ova. When deemed pertinent error propagation has also been presented in Appendix A: Additional binary isotherm datasets.

The traditional method of calculating adsorbed concentration q via a mass balance equation is associated with a random error in q which remains relatively constant across all the NaCl levels studied. When the adsorbed concentration is small, this error is a significant relative of the signal. Here an alternative method of ascertaining the adsorbed concentration via elution has been explored. It was found that the random error in adsorbed concentration calculated in this way is proportional to the measured adsorbed concentration. This is can be exploited when the adsorbed concentration is low (meaning the relative random error in q is high but low in q^*), as exploited for the BSA-Ova binary isotherm at pH 9 in 200 mM NaCl.

Another confounding factor in multicomponent isotherm study is the number of data points required to explore the competition between proteins as well as multiple NaCl levels. This is an additional reason why the lysozyme Cyt-C binary isotherm has been extensively studied as the Cyt-C chromophore facilitates rapid analytics. Here an alternative rapid analytical method for chemically similar compounds lacking chromophores has been used to explore 3 binary isotherms which exhibit different competitive behaviour. The use of UV spectra in conjunction with multivariate data analysis has facilitated the rapid study of 3 binary isotherms. Moreover, the agreement between q and q^* explored here has been good in 25/30 cases (including data sets presented in Appendix A: Additional binary isotherm datasets), building the case that this is a robust approach for the study of binary isotherms and that systematic error is under reasonable control.

Comparison in experimental time taken to generate binary isotherms using more traditional HPLC methods vs UV spectra and PLS models is shown in Figure 6-25. The analysis time displayed in Figure 9 is shorter by a factor of 7 compared to a fast HPLC method using a monolith column which is amenable with higher flow rates and shorter run times (Viklund, Svec, Frechet, & Irgum, 1997). Reduction of the analytical bottle neck by using UV spectra in conjunction with PLS models not only facilitates the generation of multicomponent isotherm data but also the analysis of multiple versions of the same data point i.e. the equilibrium as well as elution samples in addition to multiple dilutions of the same sample improving the reliability of the data generated. Assuming overnight analysis using a fully automated ALHS and with integrated plate reader the output of one PreDictor plate can be analysed at equilibrium and elution at 3 dilution levels each before the next working day, which is not possible using HPLC with short run times. This allows a greater number of both experimental and analytical replicates to be assayed, which increases the robustness of the data and their availability makes outlier detection/rejection feasible. This would not be realistic using slower conventional HPLC methods. Additionally this method of analysis is non-destructive meaning the same samples could be carried forward for further analysis if required.

In the next chapter the methodologies explored here will be extended to ternary isotherms for which there is barely a handful of examples in the literature. The PLS model performance for ternary mixtures discussed in Chapter 5, was not as convincing as for the binary cases explored, meaning their study will be more challenging.

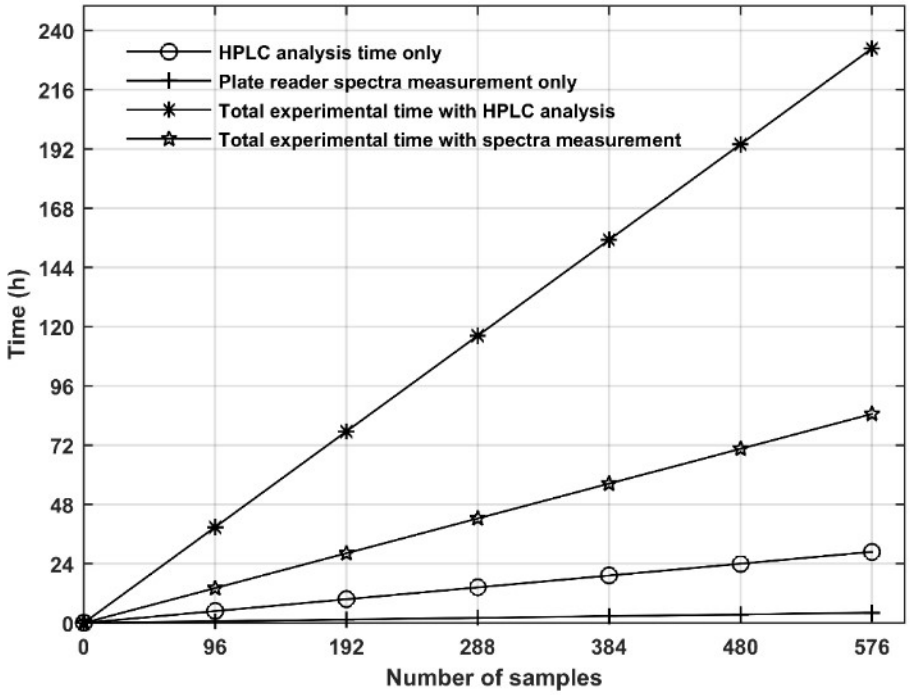

Figure 6-25 Comparison of analysis time of samples using HPLC versus UV spectra on plate reader. HPLC analysis run time was taken as 3 minutes (Viklund et al., 1997). *For UV spectra on plate reader the run time is 26 seconds for one sample. For total experiment time it was assumed that 3 analytical replicates of equilibrium samples and 3 replicates of elution samples of each plate was run in addition to time taken to run the slurry plate and prepare samples.*

Chapter 7 Ternary isotherms and their associated error

7.1 Aims and objectives of chapter

In Chapter 4 a platform methodology was developed and presented for the generation of single component adsorption isotherms using high-throughput (HT) methods and automated liquid handling systems (ALHS). In Chapter 4 quantification of adsorbed concentration in isotherms was explored using 2 methods: Firstly the traditional method; here the starting liquid concentration is known and the ending concentration is assayed, the adsorbed concentration is then calculated via mass balance and is referred to as q. The alternative method is to elute what was adsorbed to the resin and assay it creating an orthogonal measure of adsorbed concentration referred to as q^*. In Chapter 5 a rapid analytical method for the quantification of individual protein titres from binary and ternary mixtures using UV spectra and multivariate data analysis was presented and discussed. In Chapter 6 the above methodologies were applied to the study of 3 binary isotherms at a range of pH and NaCl levels. The systematic and random error in these binary isotherms was also explored.

The general aims of this chapter are to present ternary adsorption isotherms for BSA-ovalbumin (Ova)-conalbumin (Con) mixtures, on the strong anion exchanger Capto™ Q at pHs 9 and 3 NaCl levels. Additionally the propagation of both systematic and random errors in q and q^* for these multicomponent isotherms will be explored. The specific aims of the chapter are:

- Present ternary isotherms of BSA, Ova and Con on Capto™ Q at pH 9 at 0 mM. 25 and 50 mM datasets are presented in Appendix B: Additional ternary isotherm datasets.
- Calculate the adsorbed concentration via the two methods discussed above
- Present the propagated systematic and random error in adsorbed concentration when using either equilibrium or elution methods

The application of methodologies discussed previously are applied it to the study of ternary isotherms. The performance of those methodologies based on error propagation will be discussed as well as some of the underlying challenges involved in the study of ternary isotherms generally as well as limitations in the methodologies explored here.

7.2 Ternary isotherms and the propagation of their error

Ternary isotherms will be presented as a series of single component isotherms at different starting concentration of competitors. For brevity error propagation of the ternaries will only be discussed when the isotherms suggest errors are significantly contributing to the data. Three ternary isotherms have been studied in this thesis, all at pH 9 at 3 different concentrations of NaCl. There is such a large quantity of data to discuss that each ternary isotherm discussion has been split into three parts, one for each protein and the 25 and 50 mM NaCl isotherms are presented Appendix B: Additional ternary isotherm datasets.

As mentioned previously the adsorbed concentration of all isotherms was calculated using 2 methods. Firstly via the traditional method where the equilibrium concentration is assayed and the adsorbed concentration is calculated via mass balance, q as displayed in equation 3-5. The new alternative method is to elute the adsorbed protein and assay the eluted fractions in order to estimate the adsorbed concentration which is represented in equation 3-7 by q^*.

The propagation of systematic and random errors has been approached in the same way as discussed in Chapter 6. Much effort has gone in Chapter 4 to reduce sources of systematic and random error, as such, the only source of systematic error has been assumed as in the quantification of individual protein concentrations from the mixture using the UV spectra-partial lease squares (PLS) method. Building the ternary PLS model for protein quantification involved the creation of many control samples of known concentration in a well populated space filling layout in 3 component space. It is possible to estimate the concentration of every known control

sample giving an estimate of the performance of the PLS model at any given control sample, when this information is used in conjunction with the Matlab function "griddata", which can interpolate values between points, it is possible to estimate the systematic error in concentration estimate at any point in the assayable space. These errors in estimated liquid concentration can then be propagated through to systematic errors in both q and q^*.

For the q estimate in adsorbed concentration the only source of systematic error is in the estimation of equilibrium concentration, as described in equation 3-5. For the q^* estimate in adsorbed concentration elution, as described in equation 3-7, there are more sources of systematic error: All the wash/elution fractions analysed using the PLS method can introduce systematic error. That systematic error can be either positive or negative, increasing or decreased the estimated concentration. The systematic errors in q^* calculation were allowed to cancel, meaning that if elution fraction 1 was associated with an overestimate and fraction 2 an underestimate these errors were allowed to cancel one another out.

There are multiple sources of random error. An estimate of these errors has been gleaned either from in house experimental data or taken from the literature and the equations used to propagate the error are described in Table 3-1 which were used in conjunction with equations 3-5 and 3-7. In order to propagate the random error in the estimation of concentrations within a mixture a relationship between estimated concentration and random error needs to be established. This was done in the same manner as discussed in Chapter 6: The control samples used in PLS model were prepared in triplicate but averaged before model building commenced. Those triplicate spectra were supplied to the ternary PLS model and concentration estimates were obtained. From those estimates average and standard deviations were calculated. In order to establish the relationship between concentration estimate and BSA standard deviation the 3 averaged estimates (one for BSA, Ova and Con) were used along with the standard deviation of BSA to perform a linear regression with averaged estimates of BSA, Ova and Con as the independent variables and standard deviation of BSA as the dependant variables. This was repeated twice

more with Ova and Con standard deviation as dependant variables. In all cases the p-values for regression for all coefficients including the intercept was < 0.05 and so all coefficients accepted as statistically significant.

$$q = (C_0 - C_{eq}) \times F \qquad 3\text{-}5$$

$$q^* = \frac{(\sum_{i=1}^{n} m_{w,i} - m_{rw,i} + \sum_{i=1}^{n} m_{e,i} - \sum_{i=1}^{n-1} m_{re,i} - m_{req})}{V_{resin}} \qquad 3\text{-}7$$

7.2.1 0 mM NaCl

7.2.1.1 BSA part of isotherm

The isotherm plotted as a series of single component isotherms at different starting concentrations of Ova and Con are displayed in Figure 7-1, the top row of 6 sub plots displays the BSA data, both *q* and *q** are both plotted in Figure 7-1. One undesirable feature of the BSA section of the isotherm is that the saturation capacity was not explored at the highest starting concentration of Con, there are experimental restrictions which means this cannot be explored with confidence, which are as follows: In Chapter 4 it was discussed that the maximum protein stock concentration should be limited to 20 mg/mL, this was to protect the liquid handling capability of the Tecan. It was also discussed that the minimum volume of resin to be used in the PreDictor™ plate is 20 µL, this is to ensure the accuracy of the isotherms as systemic errors appear in moderately interacting conditions when 2 µL of resin is used, (there is no volume between 2 and 20 µL commercially available for Capto™ Q). A consequence of this is that the combination of protein loads available during study is capped to 500 mg/mL resin, for the single component this presents no issue as the maximum saturation capacity observed is ≈ 200 mg/mL. For the binaries the load of a single protein when the mixture is 1:1 is limited to 250 mg/mL of resin, for a single protein, again this does not present a significant issue as the maximum saturation capacity observed for the resin is 200 mg/mL. In the case of the ternary for a 1:1:1 mixture the maximum load possible is 167 mg/mL of resin which is below the maximum saturation capacity observed. This would not be problematic if all 3 proteins supressed one another's binding capacity equally because then the each of

the components would have received a large enough load to explore their saturation capacity. In reality, and as presented in Figure 7-1, it was observed that the dominant protein (BSA) binds to the resin displacing the weakest protein (Con), as the saturation capacity has not been significantly reduced for BSA the plateau of the isotherm cannot be explored at the highest combination loads of Con. This can be seen in Figure 7-1 for the BSA isotherm when the Con starting concentration is 10.4 mg/mL and Ova is at 1.4 mg/mL, only part of the linear section of the BSA isotherm can be explored.

Firstly, the agreement between q and q^* is good when the starting concentration of Ova is less than or equal to 5 mg/mL, above that the disagreement starts to become significant. The starting concentration of Con does not seem to significantly affect q and q^* agreement when considering the adsorbed BSA concentration. As q /q^* agreement is often poor, propagation of systematic error in adsorbed concentration both for q and q* will be discussed.

7.2.1.2 Propagation of systematic errors

Figure 7-2 shows adsorbed concentration of BSA as a series of contour plots. Each contour sub-plot can be considered a binary isotherm of BSA-Ova and the different subplots represent different starting concentrations of Con. This has been plotted as a reference to facilitate the discussion of propagated errors which are plotted in the same manner. Relative systematic error in q is shown in Figure 7-3 across all the sub-plots the systematic error remains largely within 0.05 or 5% which is acceptable. Looking at the relative systematic error in q^* displayed in Figure 7-4 systematic error often exceed 0.05 or 5%. There is a blank patch in the contour plot as an elution data point has been omitted. This is because q^* was in such poor agreement with q that the outlier rejection omitted q^* data for that protein load. The areas in which this 5% mark is exceeded corroborate with what was observed in Figure 7-1 which is that in general the higher Ova levels show increased relative systematic errors, therefore only the q data was used for this dataset and is displayed in Figure 7-11.

7.2.1.3 BSA part of isotherm displaying *q* data only

Figure 7-11 shows the result of omitting the *q** data and only plotting the *q* data in the top row of 6 plots. Clearly the BSA is supressed by Ova quite effectively, as the Ova starting concentration increases the plateau and slope of the linear section of the isotherm drops. This trend of effective displacement of BSA by Ova continues as the starting concentration for Con increases. Con however is very poor at displacing BSA with the slope and saturation capacity barely being affected by increasing Con starting concentration. This observation that Ova effectively displaces BSA but Con is ineffective at displacement was also observed in the binary isotherms.

y

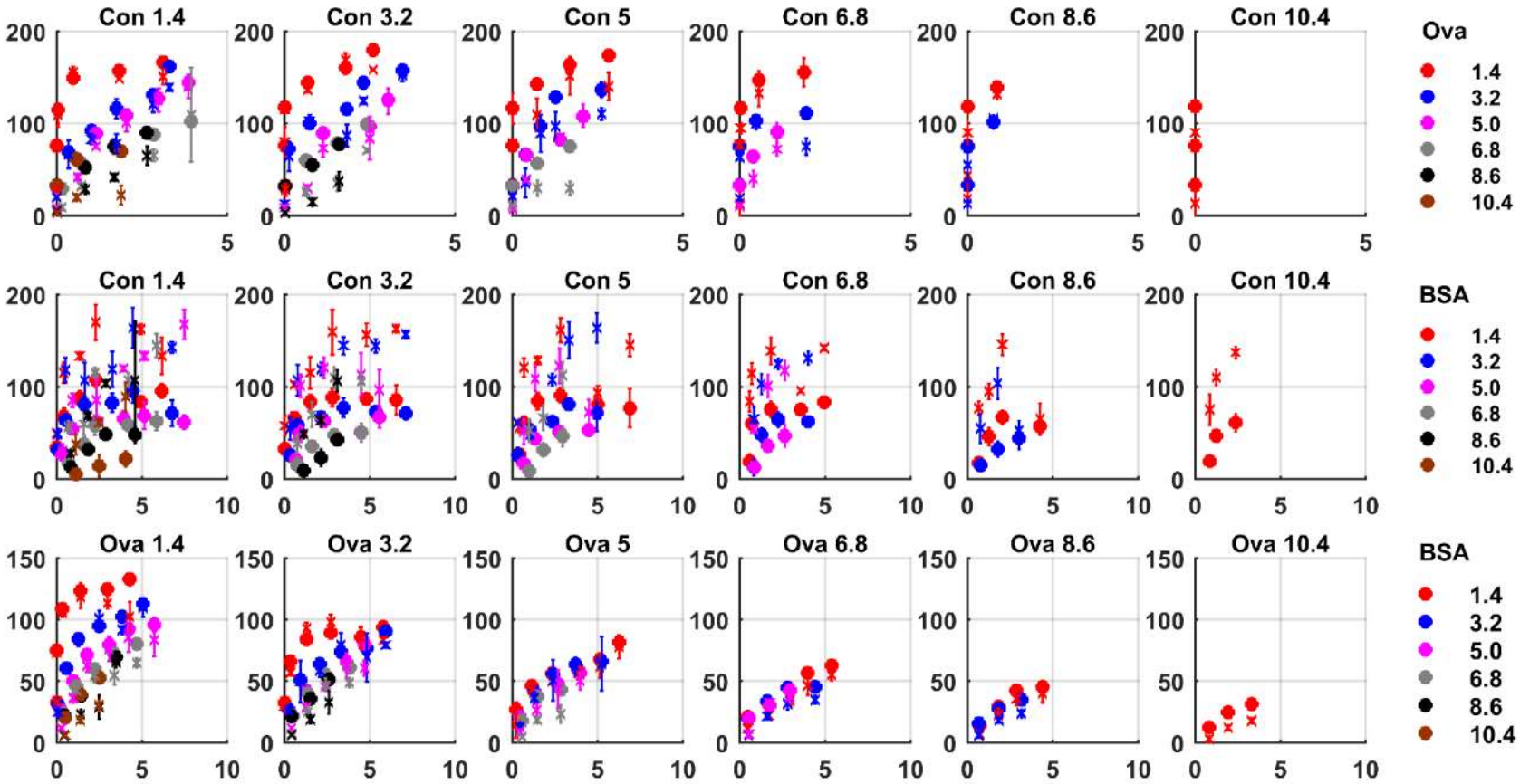

Figure 7-1 Ternary isotherm of BSA, Ova and Con plot as a series of single component isotherms at different starting concentrations of competitor protein at pH 9 and 0 mM NaCl. The plot titles display the starting concentration of one of the competitors and the legends display the starting concentration of the second competing protein in mg/mL. All x-axis display liquid concentrations in mg/mL and all y-axis represent adsorbed concentration in mg/mL of resin. Top row of 6 plots display BSA isotherms, second row of 6 plots displays Ova isotherms and the third row of 6 plots displays Con isotherms. Circles represent q data and x represents q. Error bars represent standard deviation of experimental data.*

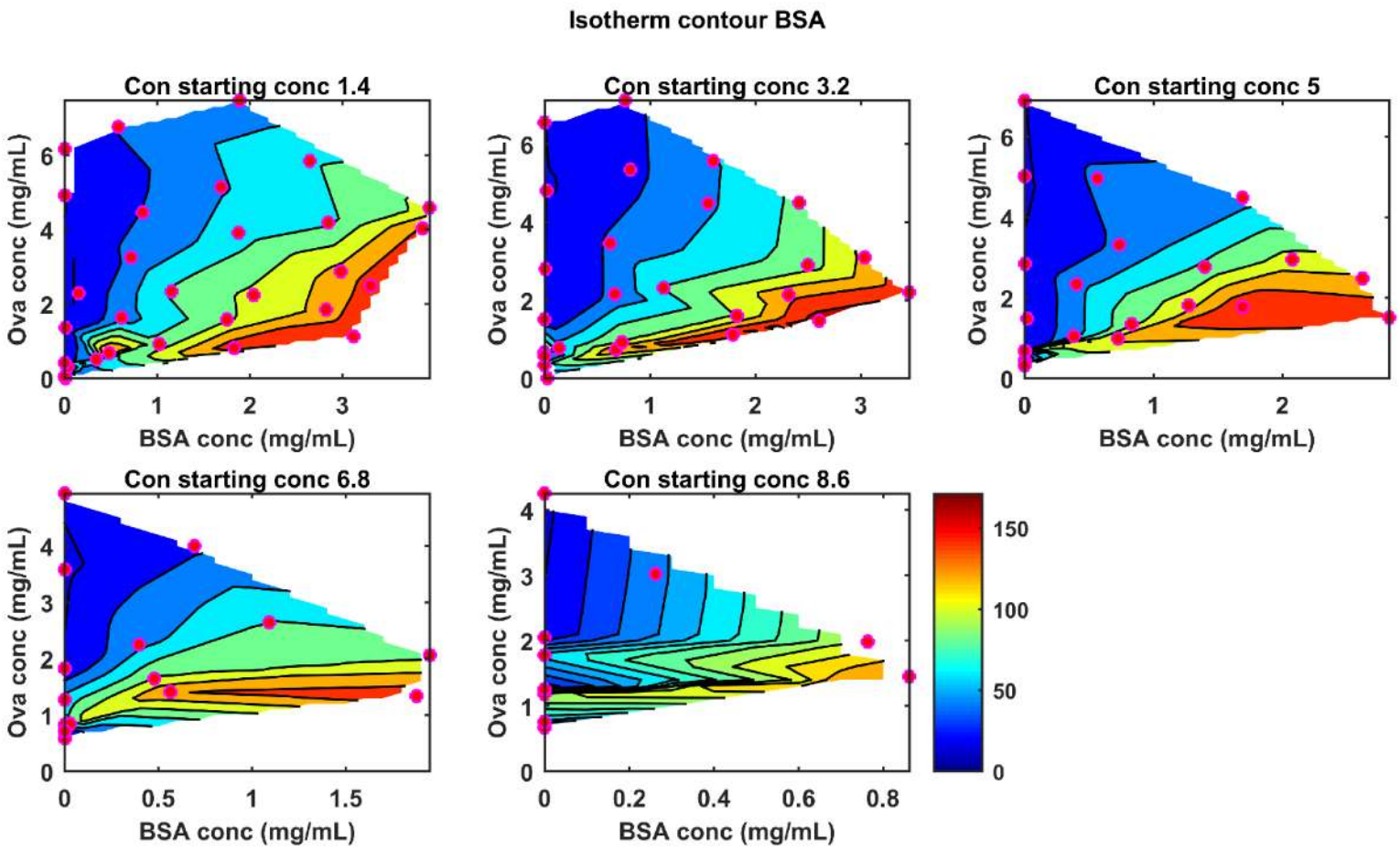

Figure 7-2 Ternary isotherm at pH 9 0 mM NaCl, contours display adsorbed concentration of BSA, adsorbed concentrations is as an average q and q. Isotherm has been drawn as a contour plot as a reference for error propagation plots. Each contour sub-plot can be thought of as a BSA-Ova binary isotherm at different starting concentrations of Con, the starting concentration of Con in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of BSA and Ova measured directly after outlier rejection.*

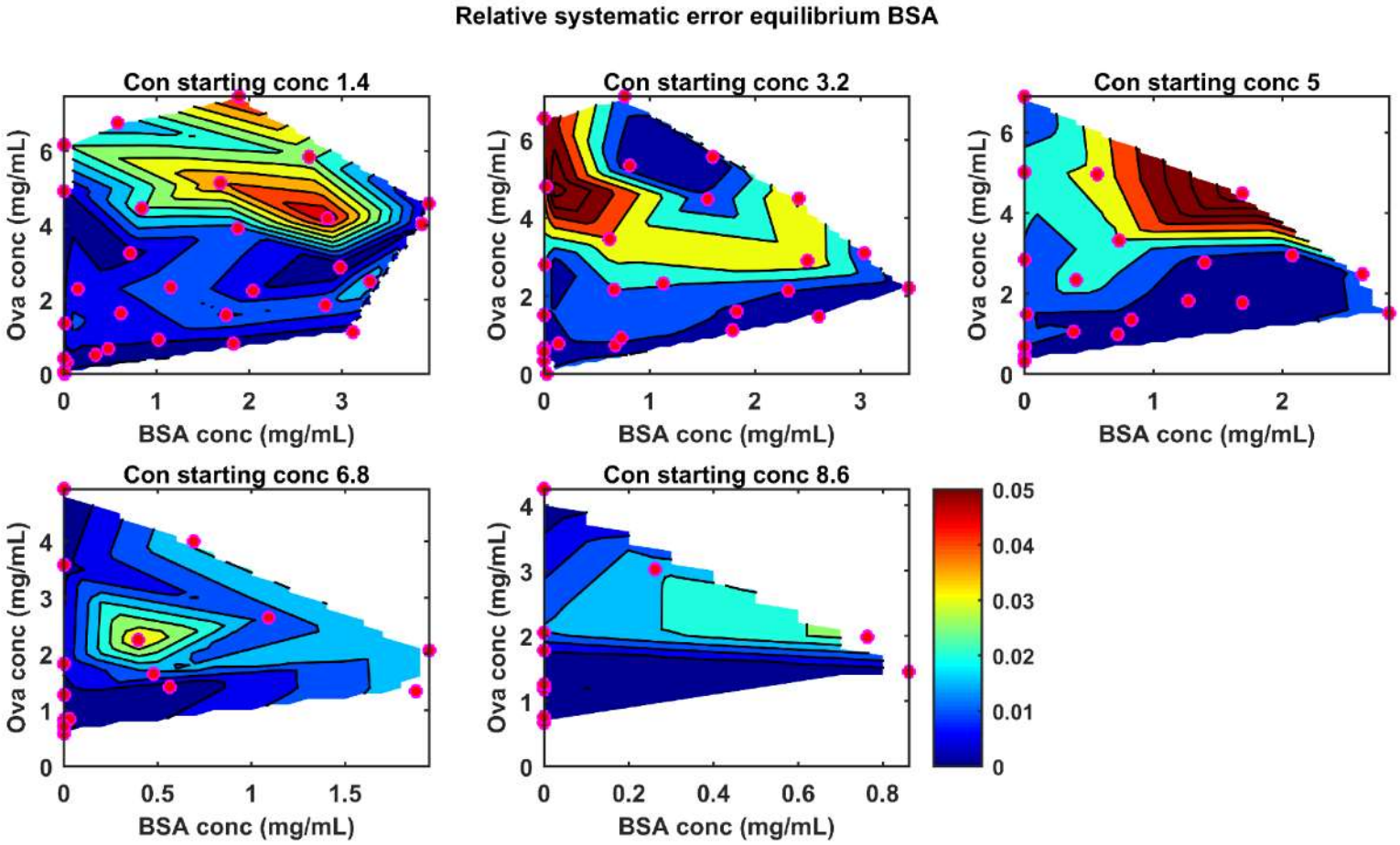

Figure 7-3 Contour plots showing propagated relative systematic error in q for BSA in ternary isotherms at pH 9 0 mM NaCl. The colours represent relative adsorbed error. Each contour sub-plot can be thought of as a BSA-Ova binary isotherm at different starting concentrations of Con, the starting concentration of Con in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of BSA and Ova measured directly after outlier rejection.

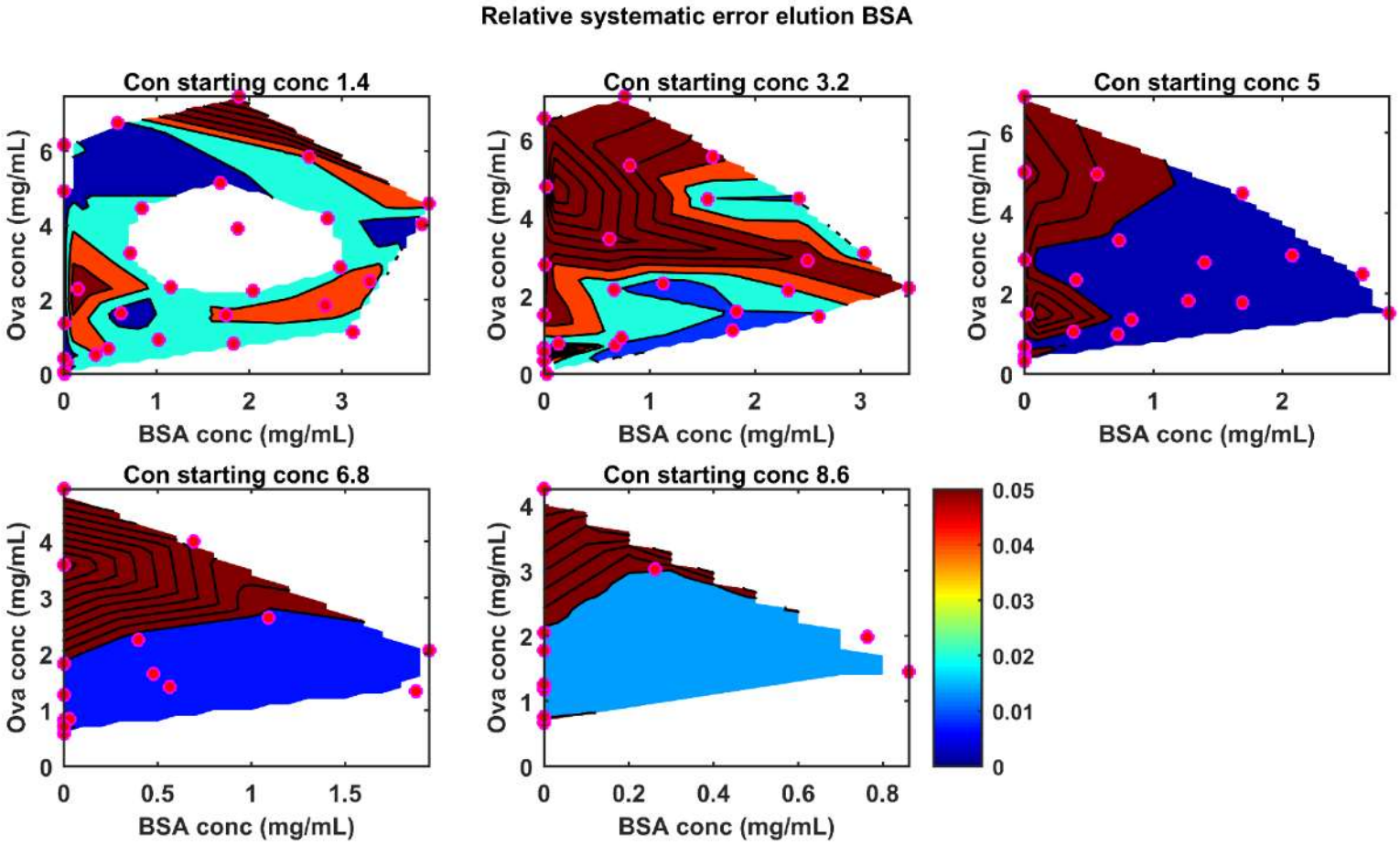

Figure 7-4 Contour plots showing propagated relative systematic error in q for BSA in ternary isotherms at pH 9 0 mM NaCl. The colours represent relative adsorbed error. Each contour sub-plot can be thought of as a BSA-Ova binary isotherm at different starting concentrations of Con, the starting concentration of Con in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of BSA and Ova measured directly after outlier rejection.*

7.2.1.4 Ova part of isotherm

Isotherm plotted as a series of single component isotherms at different starting concentrations of BSA and Con are displayed in Figure 7-1 across the second row down of sub plots, *q* and *q** are both plotted. Values of *q* and *q** almost never agree across all of the loads of competitor tested, as such propagated systematic error in *q* and *q** will be explored.

7.2.1.5 Propagation of systematic errors

Figure 7-5 shows adsorbed concentration of Ova as a series of contour plots. Each contour sub-plot can be considered a binary isotherm of BSA-Ova and the different subplots represent different starting concentrations of Con. This has been plotted as a reference to facilitate the discussion of propagated errors which are plotted in the same manner. Relative systematic error in *q* is shown in Figure 7-6 and relative systematic error in *q** in Figure 7-7. These plots show that relative errors get above 0.05 or 5% regularly when using either *q* or *q**, as such neither one measure nor the other can be used to reliably ascertain an estimate of adsorbed concentration throughout.

7.2.1.6 Ova part of isotherm after trimming and manual omission

In order to present a reasonable dataset for adsorbed concentration of Ova for the ternary the data set needs to be cleared of points which are associated with a problematic systematic error. As such the isotherm data points will be trimmed. Trimming involves removing any data points associated with systematic error greater than 5% in the assayed concentration of any of the equilibrium or elution fractions. Additionally data points are also trimmed which are associated with greater than 5% propagated error in adsorbed concentration.

The result of trimming the Ova data set is displayed in Figure 7-11. The isotherm appears much improved after trimming, much of the elution data has been omitted and trends can be more clearly observed, there is however still data present which are not plausible. There is some instances of crossover where higher loads of starting competitor have greater adsorbed concentration than lower competition

loads. This being the case the data set will be cleared more thoroughly, data points which cross over across competition loads will be manually omitted. The aim is to create a reliable subset of data points which can be used to fit to an isotherm description, the manual trimmed dataset is displayed in Figure 7-12, the Ova isotherms are displayed in the middle row of 6 plots. The isotherms show the same pattern observed for the binary isotherms, Ova is effectively displaced by BSA but Con has very little effect on the Ova isotherm.

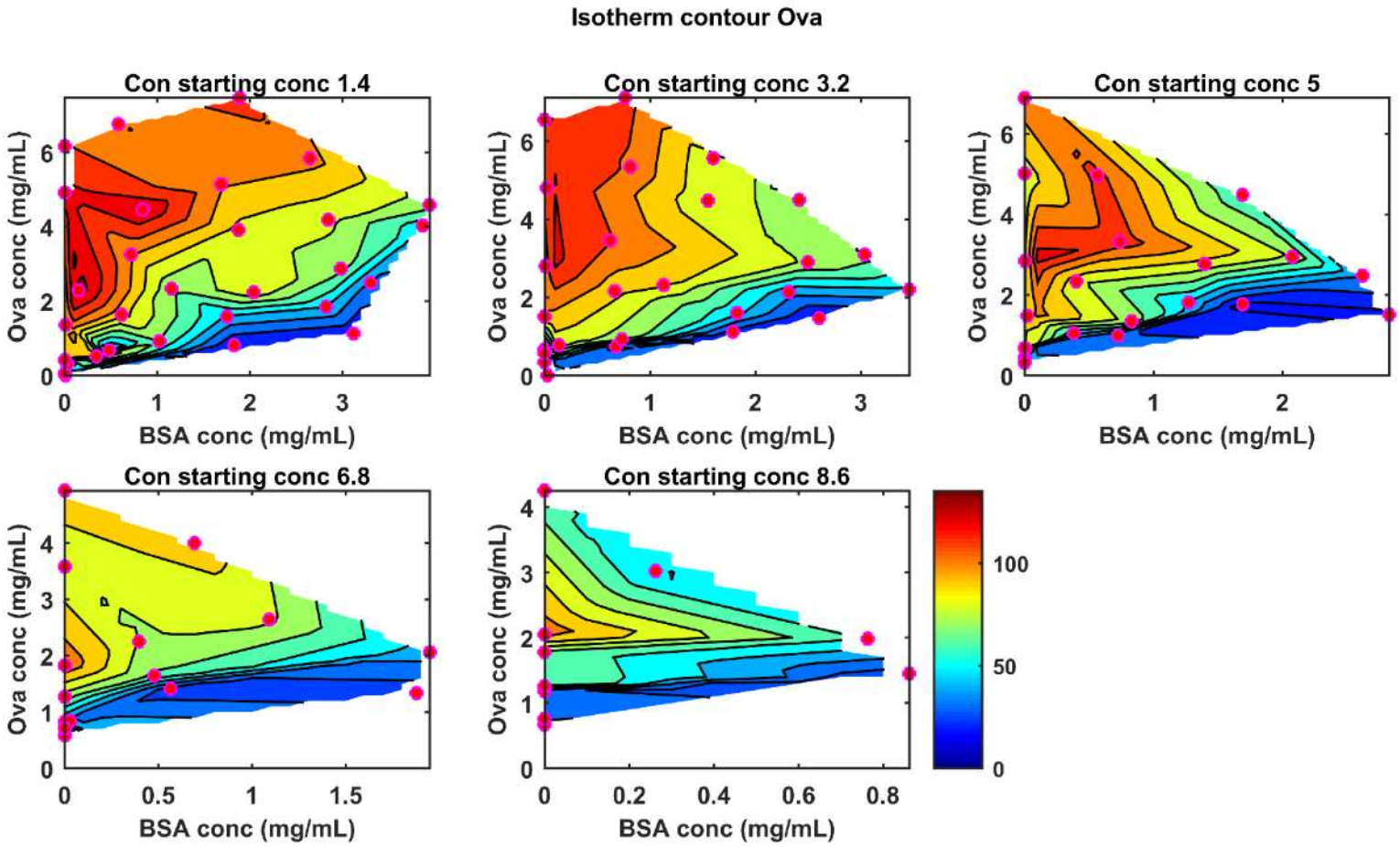

Figure 7-5 Ternary isotherm at pH 9 0 mM NaCl, contours display adsorbed concentration of Ova, adsorbed concentrations is as an average q and q. Isotherm has been drawn as a contour plot as a reference for error propagation plots. Each contour sub-plot can be thought of as a BSA-Ova binary isotherm at different starting concentrations of Con, the starting concentration of Con in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of BSA and Ova measured directly after outlier rejection.*

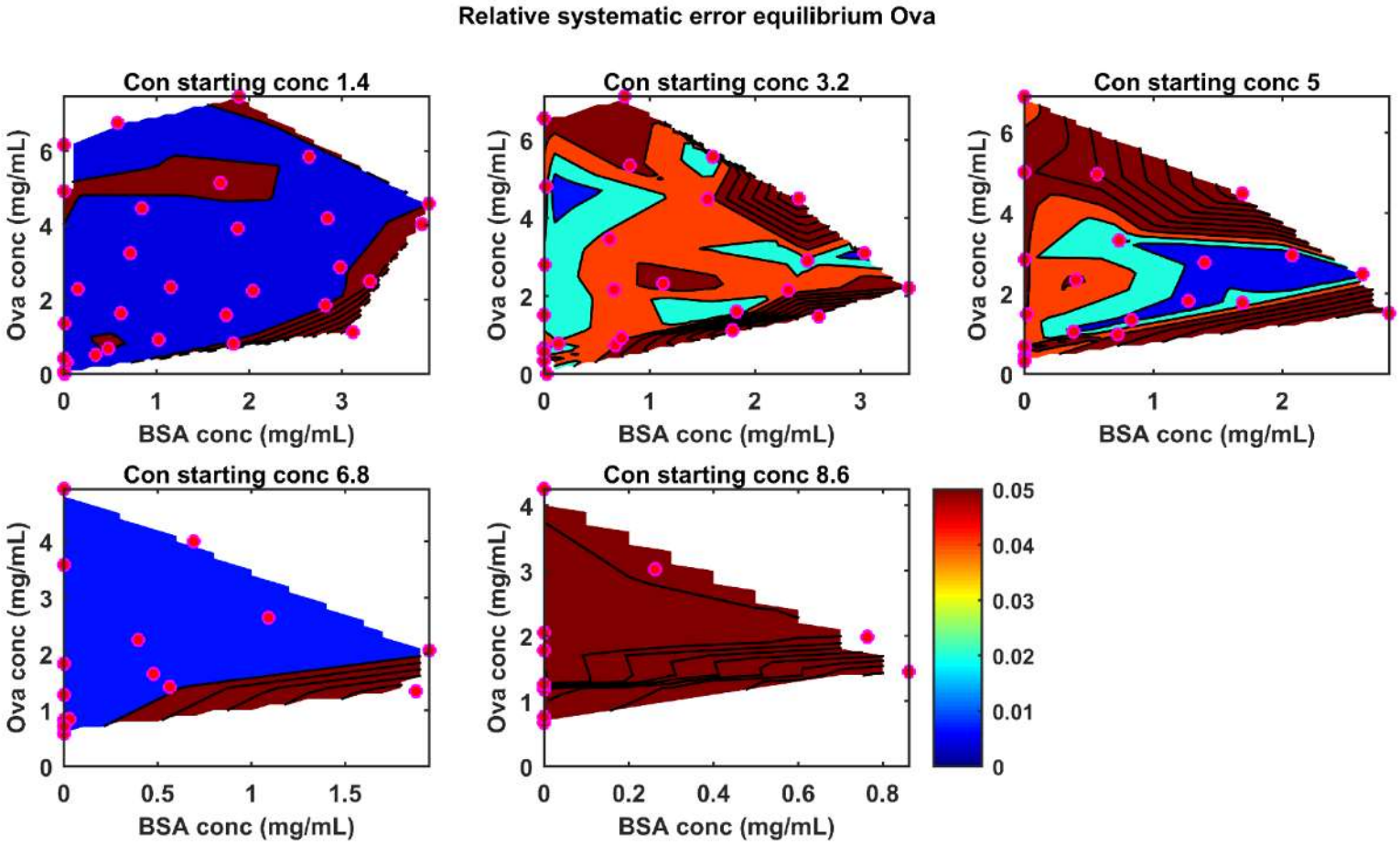

Figure 7-6 Contour plots showing propagated relative systematic error in q for Ova in ternary isotherms at pH 9 0 mM NaCl. The colours represent relative adsorbed error. Each contour sub-plot can be thought of as a BSA-Ova binary isotherm at different starting concentrations of Con, the starting concentration of Con in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of BSA and Ova measured directly after outlier rejection.

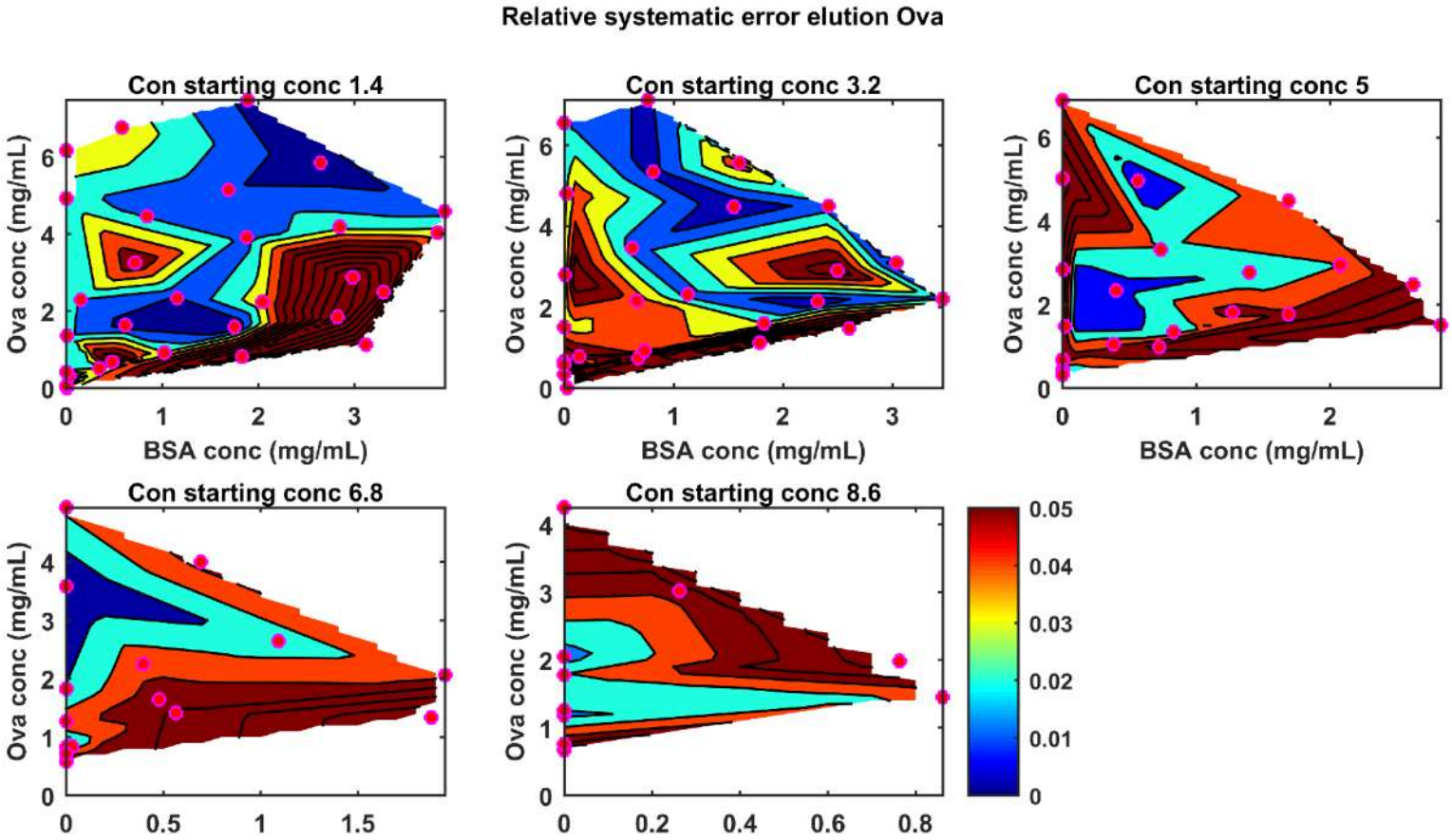

Figure 7-7 Contour plots showing propagated relative systematic error in q for Ova in ternary isotherms at pH 9 0 mM NaCl. The colours represent relative adsorbed error. Each contour sub-plot can be thought of as a BSA-Ova binary isotherm at different starting concentrations of Con, the starting concentration of Con in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of BSA and Ova measured directly after outlier rejection.*

7.2.1.7 Con part of isotherm

Con part of the ternary isotherms is displayed in Figure 7-1 across the bottom row of 6 plots. The agreement between equilibrium and elution data sets is generally quite good, particularly where starting concentration of the competitors is low. The *q** tends to underestimate the adsorbed concentration compared to *q* when the starting concentration of BSA or Ova is higher. As such the systematic error propagation of Con isotherm will be explored to determine which dataset is associated with less systematic error.

7.2.1.8 Propagation of systematic errors

Figure 7-8 displays the adsorbed concertation of Con in the ternary as a series of contour plots, each contour sub-plot can be thought of as an Ova-Con binary at a different BSA starting concentration, these plots act as a reference to compare with propagated error.

Figure 7-9 and Figure 7-10 show propagated relative systematic error in adsorbed concentration, Figure 7-9 shows error in *q* Figure 7-10 error in *q**. There is a missing part of the contour in Figure 7-10, this is because all the elution data points were omitted at this load because they agreed so poorly with the *q* data, that they were identified as outliers and omitted. Both estimates of q show systematic errors above 5% for significant parts of the isotherm, Con is the least retained species and is displaced effectively by both BSA and Ova meaning the adsorbed concentrations for Con are smallest making it more challenging to keep errors below 5% for Con. As the systematic error in both *q* and *q** exceed 5% for much of the isotherm the data set will be trimmed (as discussed in 7.2.1.1) to exclude systematic error greater than 5% in the assayed concentration of any of the equilibrium or elution fractions. Additionally data points are also trimmed which are associated with greater than 5% propagated error in adsorbed concentration. The result of this trimming is displayed in Figure 7-11.

7.2.1.9 Con part of isotherm after data has been trimmed

Figure 7-11 shows very good agreement between remaining q and q^* data points, as such trends in the isotherm will be discussed. Con is effectively displaced by both BSA and Ova with suppression in the saturation capacity and rapidly decreasing rectangularity observed as the competing starting protein concentrations increase. This is the same behaviour that was observed for the binary Con isotherms.

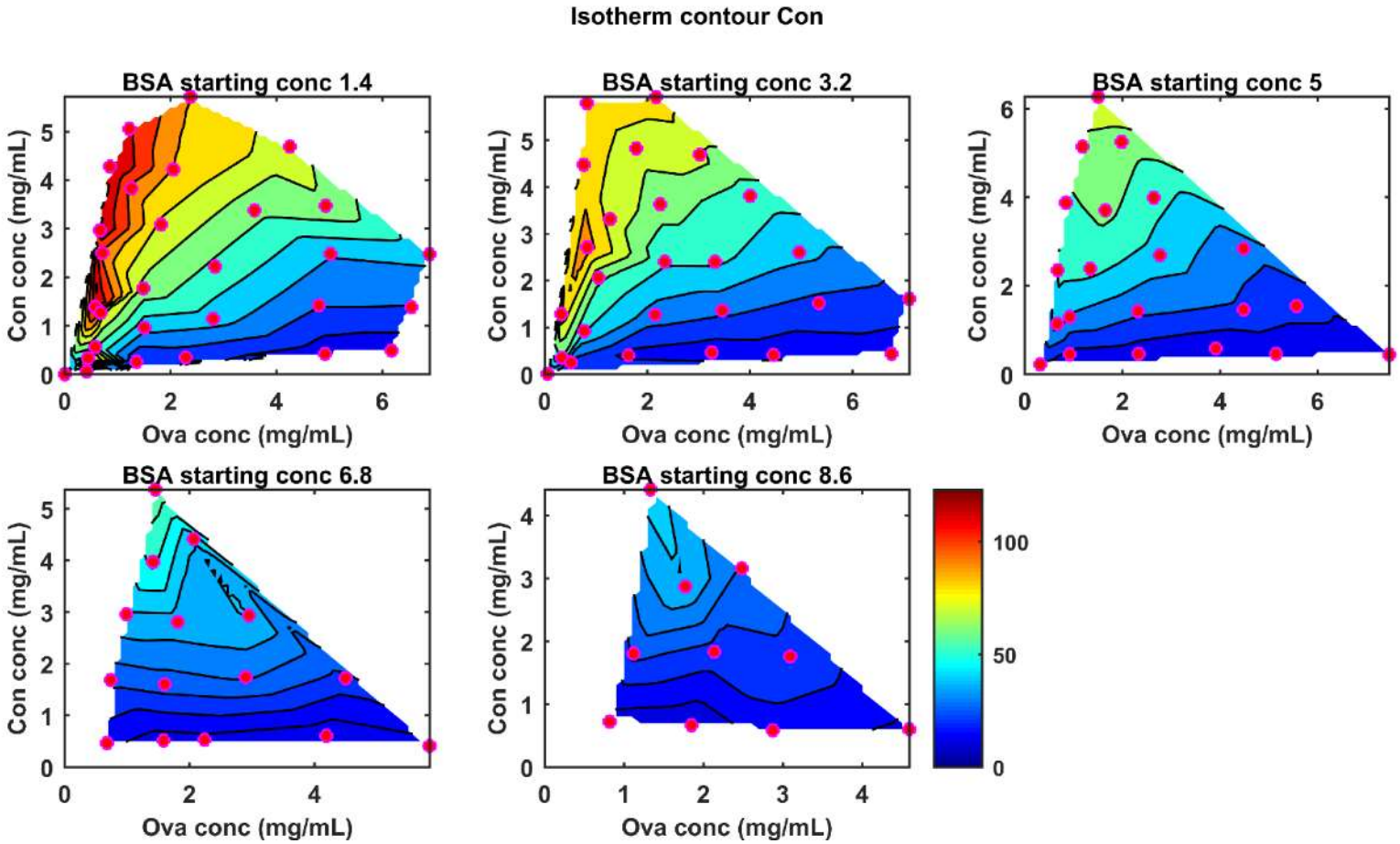

Figure 7-8 Ternary isotherm at pH 9 0 mM NaCl, contours display adsorbed concentration of Con, adsorbed concentrations is as an average q and q. Isotherm has been drawn as a contour plot as a reference for error propagation plots. Each contour sub-plot can be thought of as an Ova-Con binary isotherm at different starting concentrations of BSA, the starting concentration of BSA in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of Ova and Con measured directly after outlier rejection.*

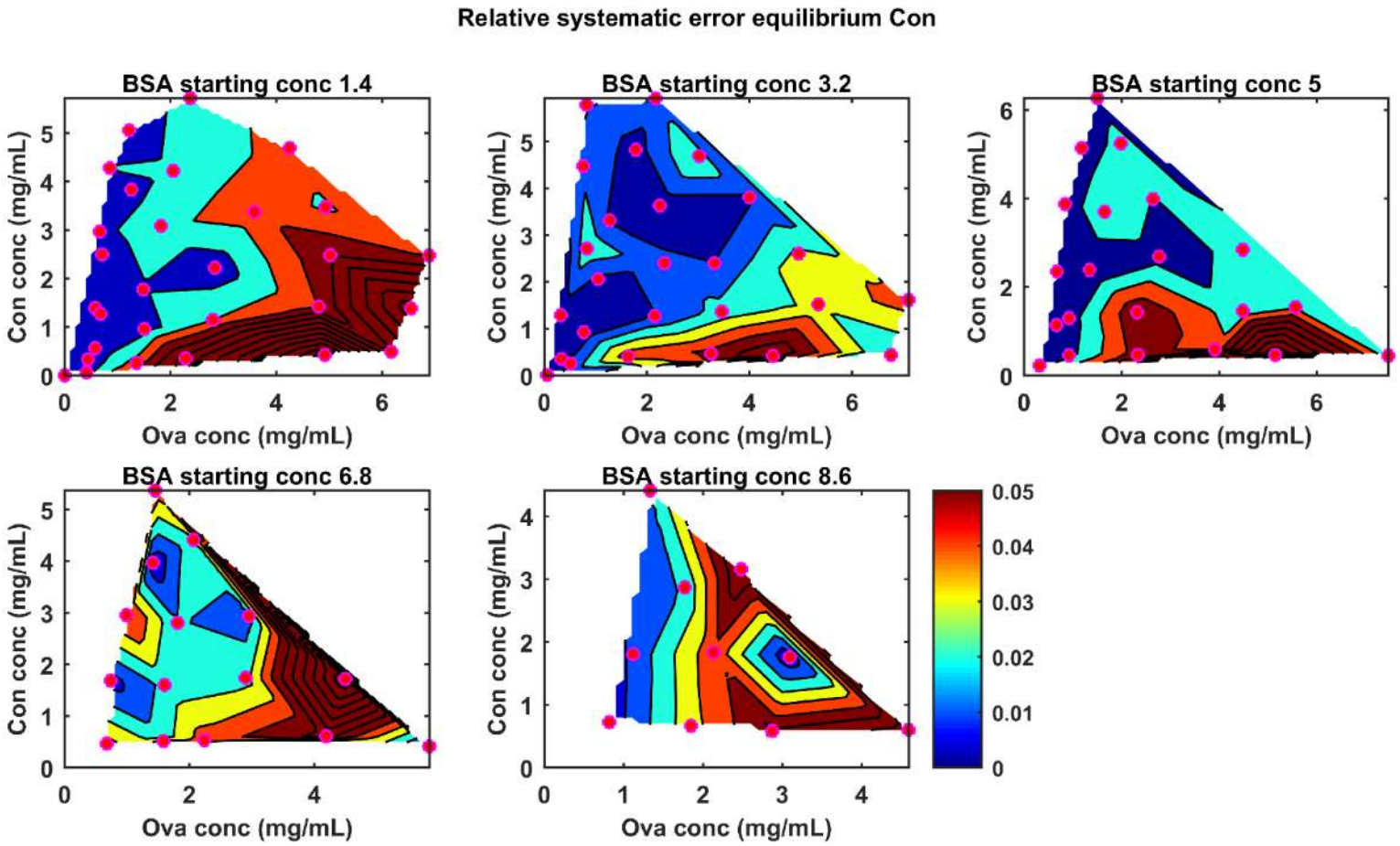

Figure 7-9 Contour plots showing propagated relative systematic error in q for Con in ternary isotherms at pH 9 0 mM NaCl. The colours represent relative adsorbed error. Each contour sub-plot can be thought of as an Ova-Con binary isotherm at different starting concentrations of BSA, the starting concentration of BSA in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of Ova and Con measured directly after outlier rejection.

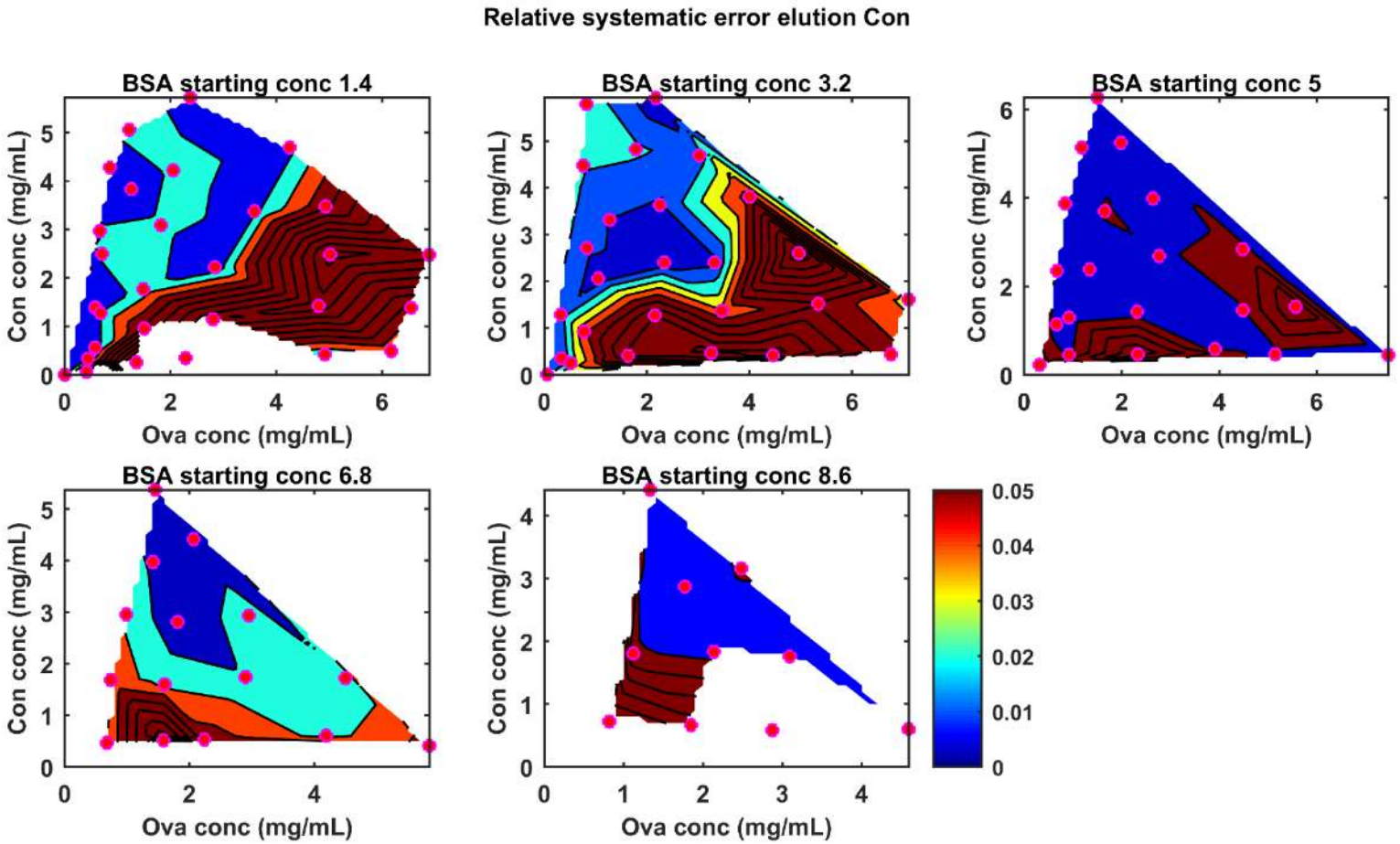

Figure 7-10 Contour plots showing propagated relative systematic error in q for Con in ternary isotherms at pH 9 0 mM NaCl. The colours represent relative adsorbed error. Each contour sub-plot can be thought of as an Ova-Con binary isotherm at different starting concentrations of BSA, the starting concentration of BSA in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of Ova and Con measured directly after outlier rejection.*

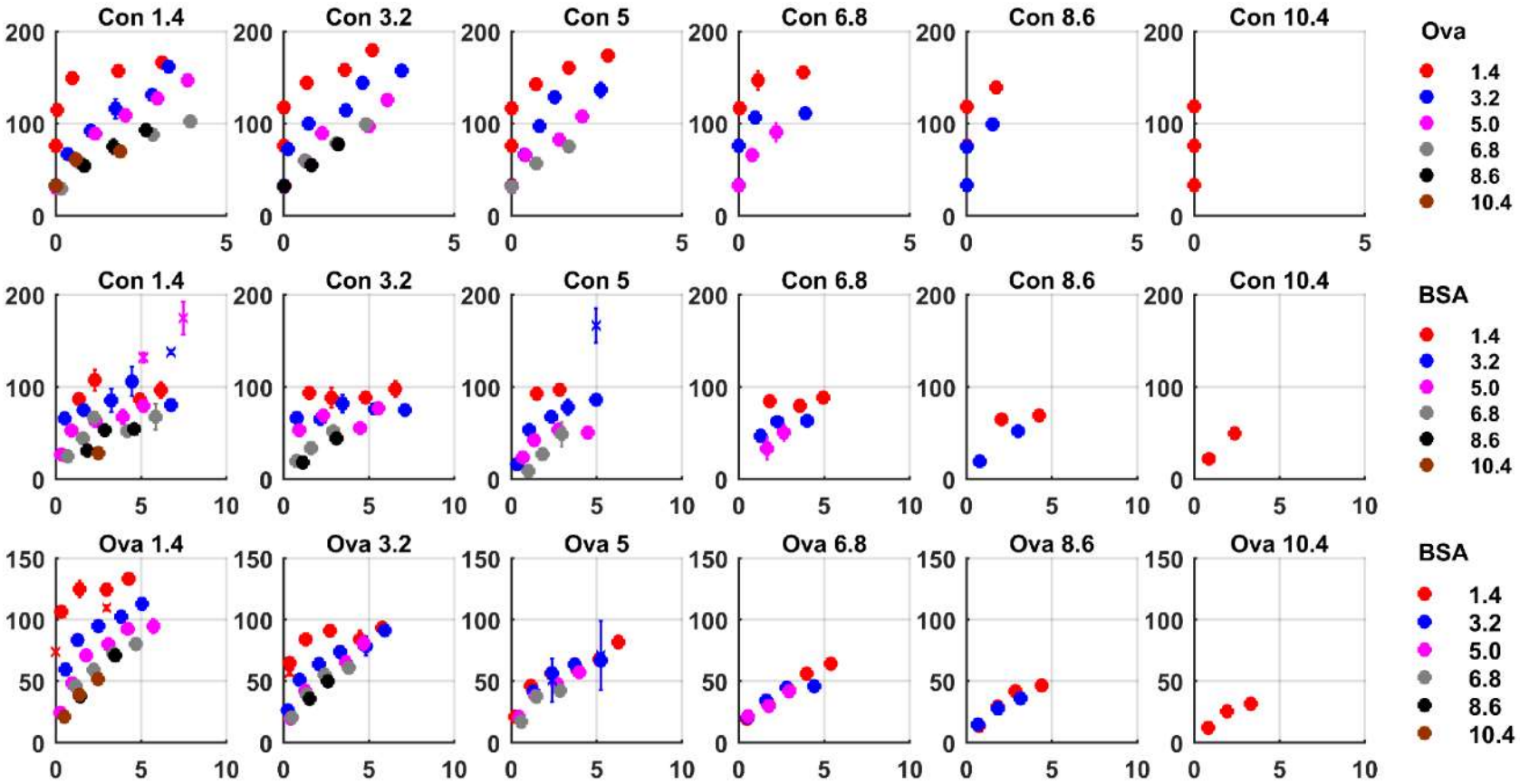

Figure 7-11 Ternary isotherm of BSA, Ova and Con plot as a series of single component isotherms at different starting concentrations of competitor protein at pH 9 and 0 mM NaCl. The plot titles display the starting concentration of one of the competitors and the legends display the starting concentration of the second competing protein in mg/mL. All x axis display liquid concentrations in mg/mL and all y-axis represent adsorbed concentration in mg/mL of resin. Top row of 6 plots display BSA isotherms, second row of 6 plots displays Ova isotherms and the third row of 6 plots displays Con isotherms. Circles represent q data and x represents q. Error bars represent standard deviation of experimental data. The BSA isotherm only displays q data, the Ova isotherm displays q and q* data after trimming and the Con isotherm displays q and q* data after trimming.*

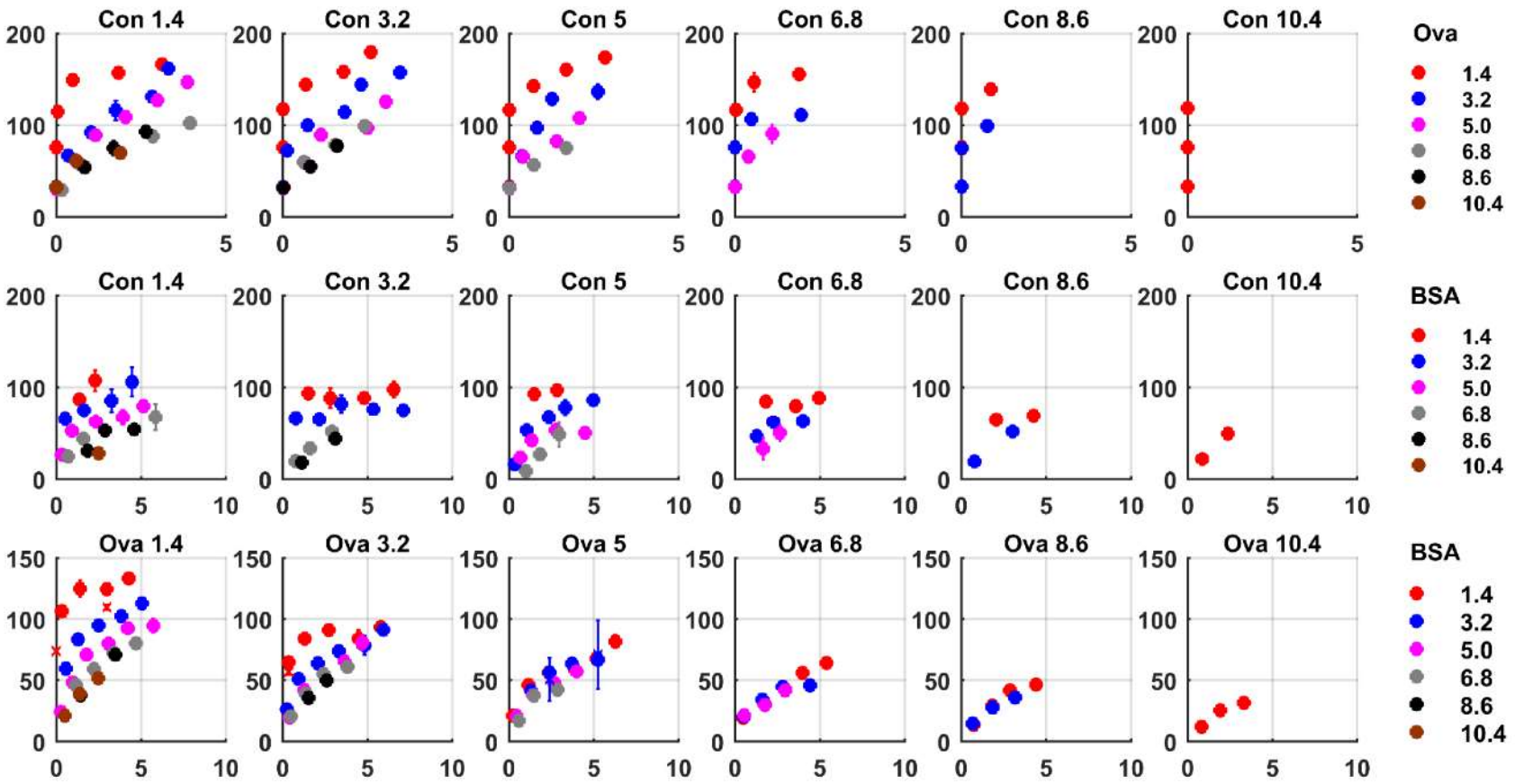

Figure 7-12 Ternary isotherm of BSA, Ova and Con plot as a series of single component isotherms at different starting concentrations of competitor protein at pH 9 and 0 mM NaCl. The plot titles display the starting concentration of one of the competitors and the legends display the starting concentration of the second competing protein in mg/mL. All x-axis display liquid concentrations in mg/mL and all y-axis represent adsorbed concentration in mg/mL of resin. Top row of 6 plots display BSA isotherms, second row of 6 plots displays Ova isotherms and the third row of 6 plots displays Con isotherms. Circles represent q data and x represent q. Error bars represent standard deviation of experimental data. The BSA isotherm only displays q data, the Ova isotherm displays q and q* data after trimming and manually omitting points, the Con isotherm displays q and q* data after trimming.*

7.3 Discussion and summary

In Chapter 6 binary isotherms were successfully quantified using the PLS method discussed in Chapter 5 after following the recommendations found in Chapter 4. For the binary isotherms there was generally good agreement between *q* and *q** with 25 out of the 30 isotherms presented suggesting there was not significant systematic error present.

The quantification of ternary isotherms presented in this chapter have proven more challenging with much poorer agreement between *q* and *q** data, the difficulty has been attributed to the increased difficulty in quantifying 3 component solutions as opposed to 2 component solutions. Chapter 5 showed that ternary mixtures were generally associated with greater accuracy error than binary mixtures. It was also shown that ternary models which spanned a smaller assayable space (0-1 mg/mL for all components) performed worse than the larger models (0-2 mg/mL for BSA and Ova and 0-3 mg/mL from Con) in assaying an external control set as discussed in section 5.2.5.2.1 Ternary performance on external data set.

As there was generally poor agreement between the *q* and *q** data. The data was assessed via the propagation of systematic errors in order to try and extract the best quality data from the overall data set. Initially both the *q* and *q** were assessed to see if either one was clearly associated with lower levels of systematic error. In cases where no data set was seen to perform significantly better, the data sets were trimmed. In this instance samples were dropped from the data set when their systematic error was found to be greater than 5% in either the directly assayed concentration or the propagated error in adsorbed concentration for both *q* and *q**. Clearer trends were observed after trimming the data sets but there were still some instances of isotherms behaving in ways which are not plausible with higher levels of competitive loads giving greater adsorbed concentrations. In these instances the data set had points omitted by manually assessing points and removing them.

Only the *q** was used for all isotherms at 50 mM NaCl which were the weakest interacting conditions. This was because the *q* data showed a tendency to return negative adsorbed concentrations. This demonstrates an advantage of using *q** data rather than *q* in conditions of weaker interaction where adsorbed concentrations are low. Subtraction of 2 numbers which are close to one another can result in a catastrophic cancelation when error is present, as in the case of the calculation of *q*. In such situations the use of *q**, which involves the addition of values, and not subject to potential catastrophic cancellation can be advantageous. This feature of *q** was also exploited in Chapter 6 in order to capture the BSA-Ova isotherm with reduced error at pH 9 200 mM NaCl.

An additional limitation of the approach was found in regard to the exploration of the binding capacity. It was difficult to explore the saturation capacity of BSA at high loads of Con. In Chapter 4 it was shown that the minimum volume of chromatography resin that should be employed is 20 µL, it was also shown that excessively high protein stock concentration creates liquid handling accuracy issues. Because there are additional restrictions on the volume of liquid that can be accommodated in a well the total protein load is somewhat restricted. In situations where one protein severely outcompetes the other the most retained species' binding capacity remains high but the load that can be explored for it is limited meaning the saturation capacity cannot be fully explored.

A recommendation from this work would be the use of self-packed slurry plates. This would allow a new set of phase ratios to be explored which could still return acceptably low systematic errors under conditions of moderate interaction but also allow higher loads to be explored in order to better saturate the resin during multi-component studies. This would still allow the use of a maximum stock concentration of 20 mg/mL which returns reasonable liquid handling performance. A sensible suggestion of resin volume would be between 10 and 15 µL. With the addition of 400 µL of liquid phase this would result in phase ratios (*F*) between 41.2 and 27.7. These phase ratios are still much smaller than 203.6 which resulted in

systematic overestimation and isotherms prone to random error. Additionally it would allow significantly higher protein loads to be explored in multi-component isotherms to allow the saturation capacity to be explored.

In Figure 7-13 a single component isotherm is displayed which has been created using the Langmuir-Freundlich isotherm with parameters which fitted the BSA isotherm well at 200 mM NaCl. This is an isotherm which interacts with moderate strength and has been considered more difficult to capture faithfully. These parameters were used to generate the same isotherm at a variety of phase ratios and resin volumes, these virtual isotherms then had error propagation applied to them. The results are displayed in Figure 7-14. A phase ratio of 20.9 using 20 µL of resin, which was has been used extensively in this thesis, results in maximum random error in q of approximately 11%. Phase ratios of 41.2 and 27.7 would result in maximum relative errors of approximately 18% and 12% respectively. This analysis suggests that 15 µL is the smallest resin volume that should be used as random errors for this more difficult isotherm are still within good control. 10 µL of resin can also produce isotherms of maximum error within 12% but only at phase ratios associated with 250 µL of liquid added which, depending on the analysis method may be approaching the lower limit of minimum assayable volume.

A second recommendation would be to explore alternative methods for the quantification of the ternary mixture. The use of UV spectra in conjunction with PLS models worked reasonably well for the elucidation of binary isotherms. However, it seems that for this ternary system the systematic errors in the method of quantification are too great to allow reliable data to be produced which can be corroborated by the comparison of q and q^* data. As was discussed in Chapter 2 The most rapid and least complex possibility would be to analyse the mixtures using analytical anion exchange chromatography Figure 2-10 and use a deconvolution method in order to quantify the overlapping regions of the peaks (Kalambet, Kozmin, Mikhailova, Nagaev, & Tikhonov, 2011). A drawback of this method is that the throughput of analysis would be decreased and the ternary study would take

significantly longer. However, the quality of the data should be prioritised over the experimental timescale or workload.

An alternative and intriguing method of multicomponent quantification would be the use of capillary electrochromatography (CEC). In this method proteins are separated by electroosmotic flow, similar to capillary zone electrophoresis. Additionally to this the capillary is coated with hydrophobic reverse phase ligands whilst an isocratic elution using acetonitrile is used to mobilise the protein within the capillary. The minimal band spreading associated with capillary separations and 2 modes of separation allows CEC to separate proteins with high separation factors than when achieved by one method alone. This method could potentially provide the baseline separation required for the model system studied here without the need for a deconvolution method (X. Huang, Zhang, & Horváth, 1999).

It should be noted that the ternary feed system used here represents a difficult separation problem. The three albumins, despite their different sources and somewhat different molecular weight, could not be separated using any analytical HPLC method researched or explored in this thesis. Together, with the lack of distinguishing chromophores, the numerical evaluation of composition from UV spectra by PLS was also difficult, and led to substantial systematic errors. The problem could therefore be thought of an analogue of a difficult separation problem of industrial relevance e.g. the resolution of product variants such as protein isoforms. The limiting factor in the elucidation of ternary isotherms is the presence

of systematic error introduced by the PLS method, and future work, should evaluate other approaches to this problem.

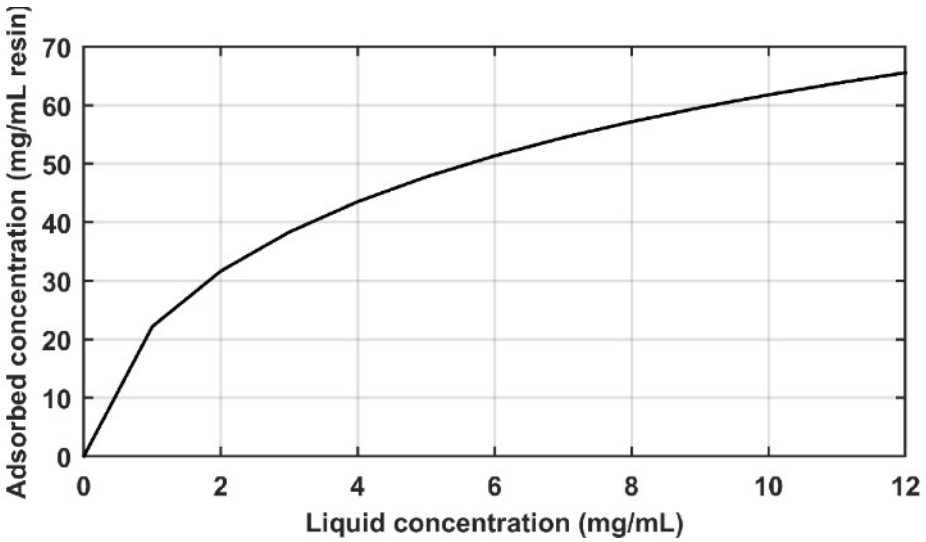

Figure 7-13 Single component isotherm generated using the Langmuir-Freundlich parameters Λ = 128, K = 0.210, α =0.650. These parameters fit the single component isotherm for BSA at pH 9 50 mM Tris and 200 mM NaCl well. These parameters were used to generate the same single component isotherm at a variety of different phase ratios in order to compare their propensity for error.

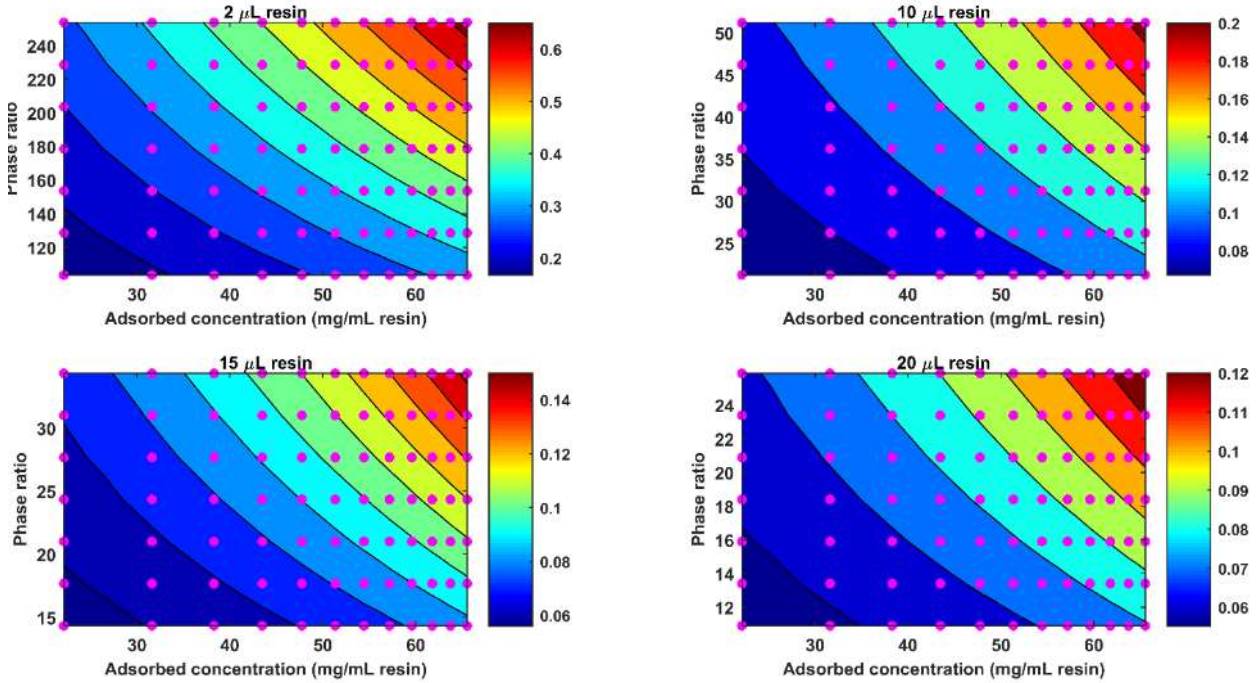

Figure 7-14 Contour plots displaying the propagated error in adsorbed concentration (q) for the single component isotherm displayed in Figure 7-13 at a range of resin volume and phase ratios (F). The volume of resin involved is shown in the title of each subplot.

Chapter 8 Fitting multicomponent isotherm descriptions

8.1 Aims and objectives of chapter

The general aim of this chapter is to fit binary and ternary isotherm data to different isotherm descriptions and assess their suitability to capture the trends seen in multicomponent isotherms. The specific aims are:

- Fit BSA-Ova, Ova-conalbumin (Con) and BSA-Con binary isotherms, and the BSA-Ova-Con ternary isotherms at pH 9 at multiple levels of NaCl to the Langmuir-Freundlich-Jovanovic (FLJ), steric mass action (SMA) and Padé isotherm descriptions.

Fitted data that which has not been discussed directly is included in Appendix C: Additional isotherm fitting data.

8.2 Multicomponent isotherm fitting

The binary isotherms at pH 9 were simultaneously fit to multiple NaCl levels using the FLJ, SMA and Padé formalisms described in equations 2-3, 2-4 8-1 and 8-2. The FLJ and SMA isotherms were discussed in Chapter 2. In the case of the multicomponent Padé isotherm for N components described in equation 8-1, q describes the adsorbed concentration, Λ is the saturation capacity, O is the order of the isotherm. All the Padé isotherms discussed here were 3rd order. In equation 8-1 C_{Cl^-} is the chloride counter ion concentration, C is the liquid equilibrium concentration of the solute and the following are all empirical constants: a, S and B. The Matlab codes involved in fitting the isotherm descriptions were written by Dr. Spyridon Konstantinidis and were run by myself.

A summary of the binary fits are displayed in Table 8-1 and the ternary fits in Table 8-2 with the R^2 values displayed. The formula to calculate R^2 is displayed in equation 8-2, y represents the dependant variable to be modelled, in this case it is adsorbed concentrations, f represents the modelled values of adsorbed

concentration and *y* with an overhead bar represents average values of adsorbed concentration for the data set. Table 8-1 and Table 8-2 display some negative values for R^2 which is explained by the sum of squared error for the average value minus the observed value being less than the sum of squared error for the modelled value minus the observed value. A negative values represents an extremely poor fit where the estimates would be better modelled by taking an average value rather than fitting the modelled value. Some of the parameters for the isotherm descriptions apply across all the NaCl levels and components involved, but the R^2 values apply either only within one NaCl level or across all the NaCl levels but for only one component which is why the average value within a sub set of data can be greater than the modelled value.

The fitting of isotherms will be broken down into different formalisms. Meaning the FLJ fits will be discussed followed by the SMA and finally the Padé.

$$q_i = \frac{K_i.\left(q_0 - \sum_{j=1}^{M}\left(z_j + \sigma_{S,j}\right).q_j\right)^{z_i}}{(C_{Cl^-})^{z_i}} C_{eq,i} \qquad 2\text{-}3$$

$$q_i = \frac{\Lambda_i.K_i.{C_{eq,i}}^{\alpha_i}.\exp(-\mathrm{A}_i.C_{Cl^-})}{1 + \sum_{j=1}^{N} K_j.{C_{eq,j}}^{\alpha_j}.\exp(-\mathrm{A}_j.C_{Cl^-})} \qquad 2\text{-}4$$

$$q_i = \frac{\Lambda_i.\sum_{o=1}^{O} a_{i,o}.C_{eq,i}^{o}.\exp(-S_{i,o}.C_{Cl^-})}{1 + \sum_{i=1}^{N}\sum_{o=1}^{O} B_{i,o}.C_{eq,i}^{o}.\exp(-S_{i,o}.C_{Cl^-})} \qquad 8\text{-}1$$

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2} \qquad 8\text{-}2$$

BSA-Ova binary

	BSA					Ova				
NaCl (mM)	0	50	100	200	Overall	0	50	100	200	Overall
FLJ	0.92	0.69	0.39	0.56	0.71	0.90	0.66	0.46	-1.02	0.69
SMA	0.70	0.54	0.21	0.70	0.54	0.66	0.48	0.38	-4.55	0.49
Padé	0.93	0.77	0.60	0.61	0.79	0.91	0.77	0.66	0.01	0.79

BSA-Con binary

	BSA					Con				
NaCl (mM)	0	25	50	-	Overall	0	25	50	-	Overall
FLJ	0.55	0.26	0.37	-	0.39	0.31	0.17	0.16	-	0.25
SMA	0.04	0.17	0.29	-	0.16	-0.66	-0.38	-0.74	-	-0.58
Padé	0.37	0.19	0.35	-	0.30	0.15	-0.57	-0.64	-	-0.17

Ova-Con binary

	Ova					Con				
NaCl (mM)	0	25	50	-	Overall	0	25	50	-	Overall
FLJ	0.85	0.86	0.72	-	0.81	0.80	0.82	0.55	-	0.76
SMA	0.73	0.49	0.57	-	0.63	-0.66	0.64	-0.81	-	-0.48
Padé	0.73	0.73	0.76	-	0.74	0.27	0.55	0.32	-	0.32

Table 8-1 Table displaying results of fitting all binary isotherm data to Freunlich-Langmuir-Jovanovich (FLJ), steric mass action (SMA) and Padé formalisms. R^2 results are displayed for the fit at each level of NaCl as well as an overall fit taking into account all the NaCl levels. The overall fits with the greatest R^2 values are highlighted in green, the worst is highlighted in red and middling values are highlighted in orange. The SMA model NaCl level was fit using NaCl concentration +9 mM as the SMA model cannot be fit to a 0 value for counter ion concentration and so the concentration of chloride ions in the buffer was added.

	BSA				*Ova*				*Con*			
NaCl (mM)	0	25	50	Overall	0	25	50	Overall	0	25	50	Overall
FLJ	0.43	0.60	0.32	0.43	0.57	0.19	0.62	0.51	0.56	0.09	-3.40	0.47
SMA	0.15	0.66	-0.20	0.14	0.46	0.28	0.54	0.45	-1.23	-0.34	-60.37	-1.74
Padé	0.12	0.66	0.14	0.27	0.62	0.34	0.77	0.62	0.00	-0.35	-2.95	-0.07

Table 8-2 Table displaying results of fitting ternary BSA-Ova-Con ternary isotherm data to Freunlich-Langmuir-Jovanovich (FLJ), steric mass action (SMA) and Padé formalisms. R^2 results are displayed for the fit at each level of NaCl as well as an overall fit taking into account all the NaCl. The overall fits with the greatest R^2 values are highlighted in green, the worst is highlighted in red and middling values are highlighted in orange. The SMA model NaCl level was fit using NaCl concentration +9 mM as the SMA model cannot be fit to a 0 value for counter ion concentration and so the concentration of chloride ions in the buffer was added.

8.2.1 Freundlich-Langmuir-Jovanovic isotherm fitting

R^2 values for the FLJ isotherm fit to all the binary isotherm data are displayed in Table 8-1. The R^2 for the overall fits using the FLJ formalism describes the Ova-Con and BSA-Ova isotherm reasonably well with R^2 values of approximately 0.7 or greater for these binary isotherms. Closer inspection of the R^2 values at each level of NaCl shows that the FLJ does not perform well at all NaCl levels. For the BSA-Ova binary the fit is poor at 100 mM NaCl and above and in the case of the Ova-Con binary the fit is poor at 50 mM NaCl for Con.

8.2.1.1 BSA-Ova binary pH 9

Evaluation of the observed and predicted values for adsorbed concentration when using the FLJ isotherm are displayed in Figure 8-1 to Figure 8-4. Figure 8-1 shows that the FLJ isotherm fits the BSA well at 0 mM NaCl, this is interesting as this isotherm has a highly rectangular shape at 1 mg/mL starting concentration of Ova. This rectangularity is often difficult to capture in isotherm formalisms. The isotherm formalism continues to fit well for BSA without significant systematic mis-estimation of adsorbed concentration except for at the 100 mM NaCl level displayed in Figure 8-2 where there is systematic overestimation of adsorbed concentration at low starting concentrations of Ova.

The observed and predicted adsorbed concentrations of Ova in the BSA-Ova binary isotherm at pH 9 at 50 mM NaCl is displayed in Figure 8-3. The FLJ isotherms show systematic mis-estimation of adsorbed concentration at NaCl levels of 50 mM and greater. Figure 8-3 shows at 50 mM NaCl the FLJ isotherm systematically underestimates the adsorbed concentration at low starting concentrations of BSA. At 100 mM NaCl the adsorbed concentration is underestimated by the FLJ isotherm at low starting concentration of BSA before crossing over and overestimating the adsorbed concentration at higher starting concentration of BSA. The prediction of adsorbed concentration of Ova at 200 mM (Figure 8-4) is extremely poor with systematic underestimation throughout.

It would seem that there is still significant systematic mis-estimation of adsorbed concentration using the FLJ isotherm to predict adsorbed concentration for the BSA-Ova isotherm at pH 9. Overall this isotherm formalism does not fully capture the isotherm trends observed.

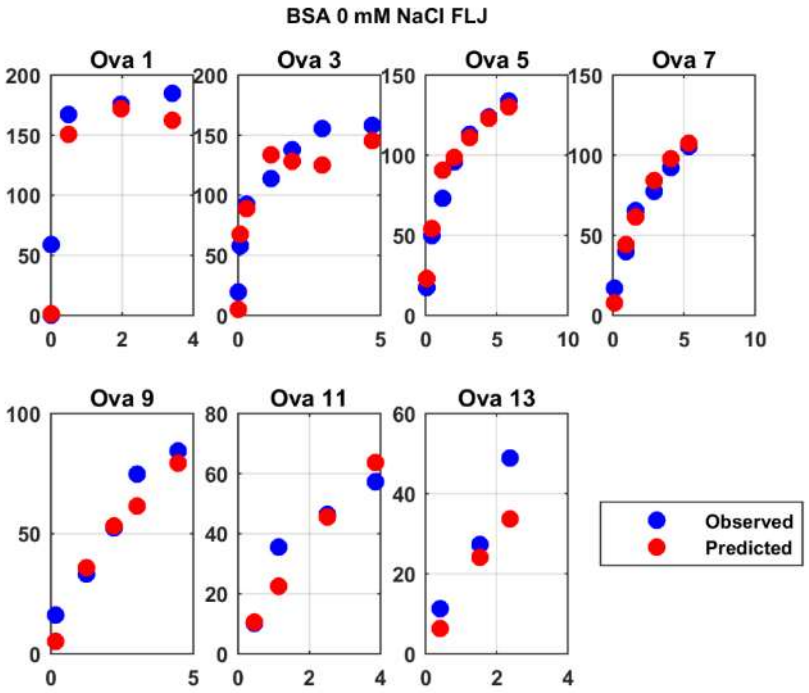

Figure 8-1 Freundlich-Langmuir-Jovanovic (FLJ) fit to binary BSA-Ova binary isotherm showing observed and predicted values for BSA adsorbed concentration at 0 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of Ova displayed as the title of each subplot.

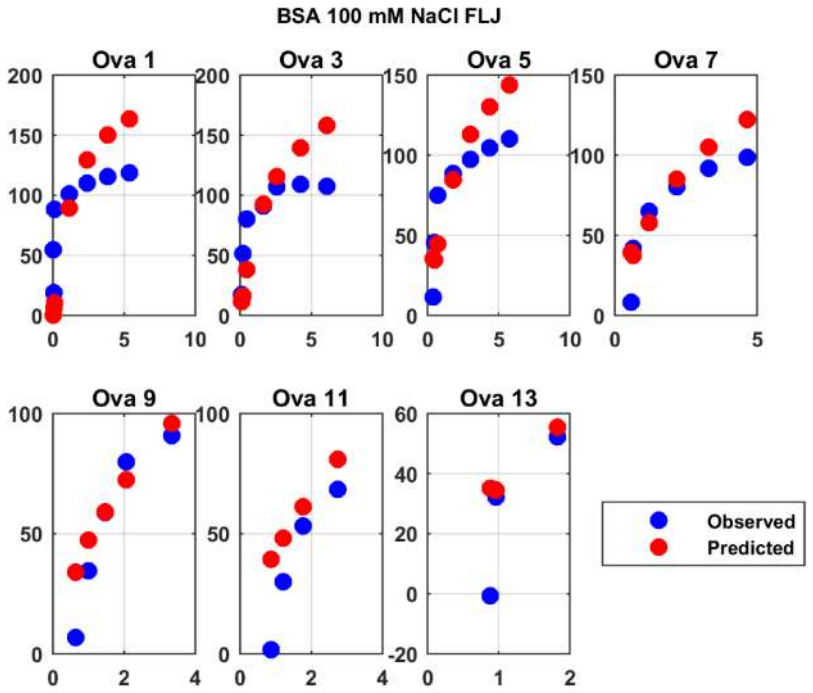

Figure 8-2 Freundlich-Langmuir-Jovanovic (FLJ) fit to binary BSA-Ova binary isotherm showing observed and predicted values for BSA adsorbed concentration at 100 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of Ova displayed as the title of each subplot.

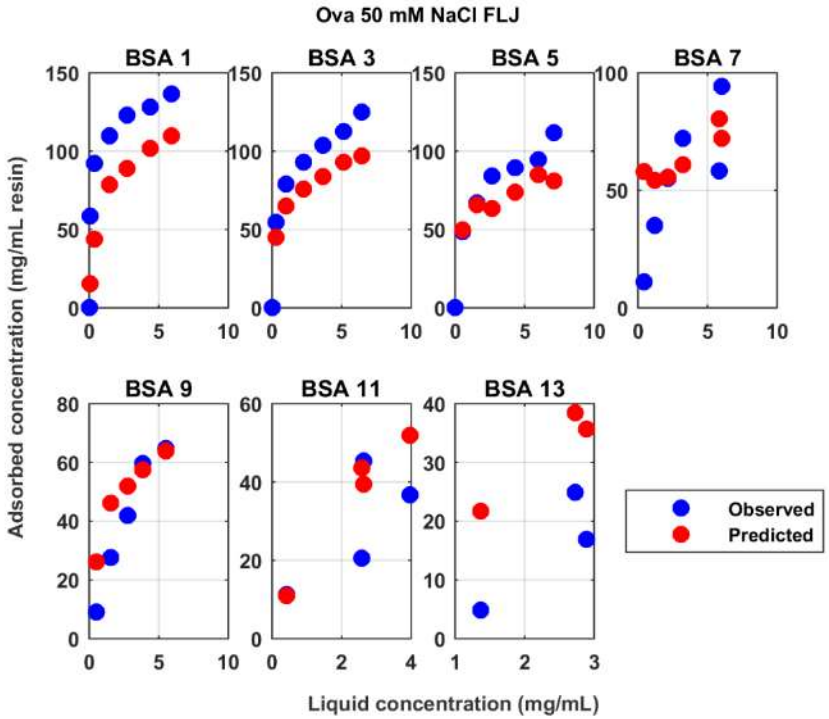

Figure 8-3 Freundlich-Langmuir-Jovanovic (FLJ) fit to binary BSA-Ova binary isotherm showing observed and predicted values for Ova adsorbed concentration at 50 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of BSA displayed as the title of each subplot.

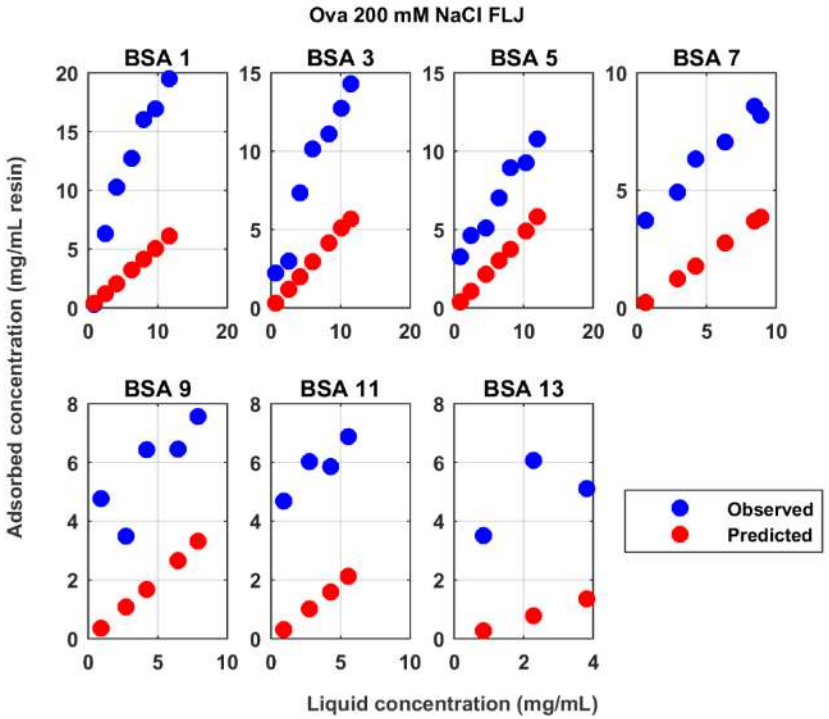

Figure 8-4 Freundlich-Langmuir-Jovanovic (FLJ) fit to binary BSA-Ova binary isotherm showing observed and predicted values for Ova adsorbed concentration at 200 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of BSA displayed as the title of each subplot.

8.2.1.2 Ova-Con binary pH 9

Comparison of the FLJ fits to the Ova-Con binary isotherm at pH 9 to multiple NaCl levels is shown in Figure 8-5 to Figure 8-7. The fit for Ova at all NaCl levels is reasonable with minimal systematic mis-estimation for Ova at all NaCl levels as shown in Figure 8-5 and Figure 8-6. Con is similarly well fit except for at the highest NaCl level of 50 mM NaCl in Figure 8-7 which shows systematic underestimation in the adsorbed concentration.

Visual assessment of the fits reflects what was shown in Table 8-1 regarding the R^2 values observed. The FLJ fits the Ova Con isotherm well except for Con at the highest NaCl levels where FLJ isotherm systematically underestimates the adsorbed concentration.

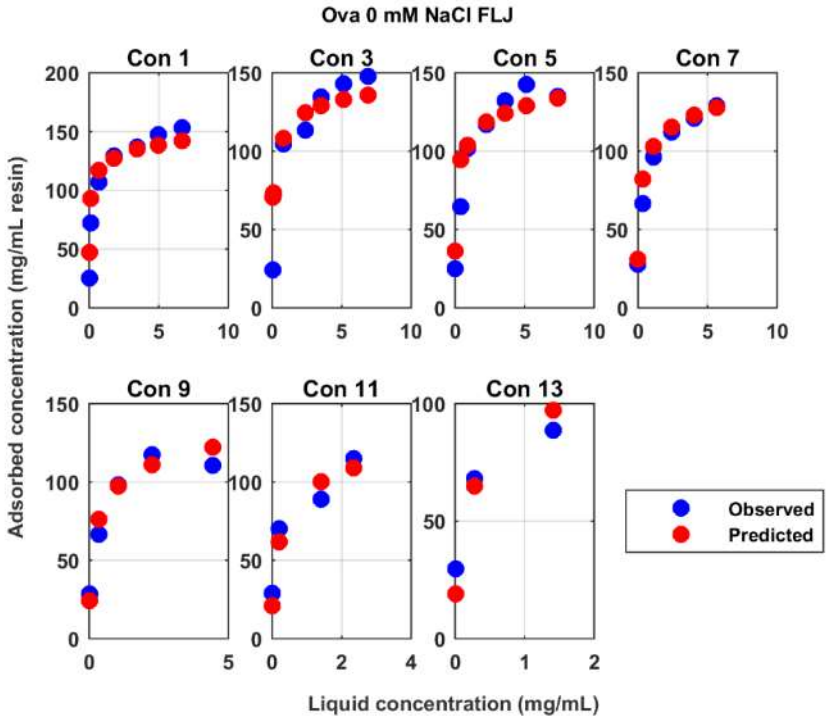

Figure 8-5 Freundlich-Langmuir-Jovanovic (FLJ) fit to binary Ova-Con binary isotherm showing observed and predicted values for Ova adsorbed concentration at 0 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of Con displayed as the title of each subplot.

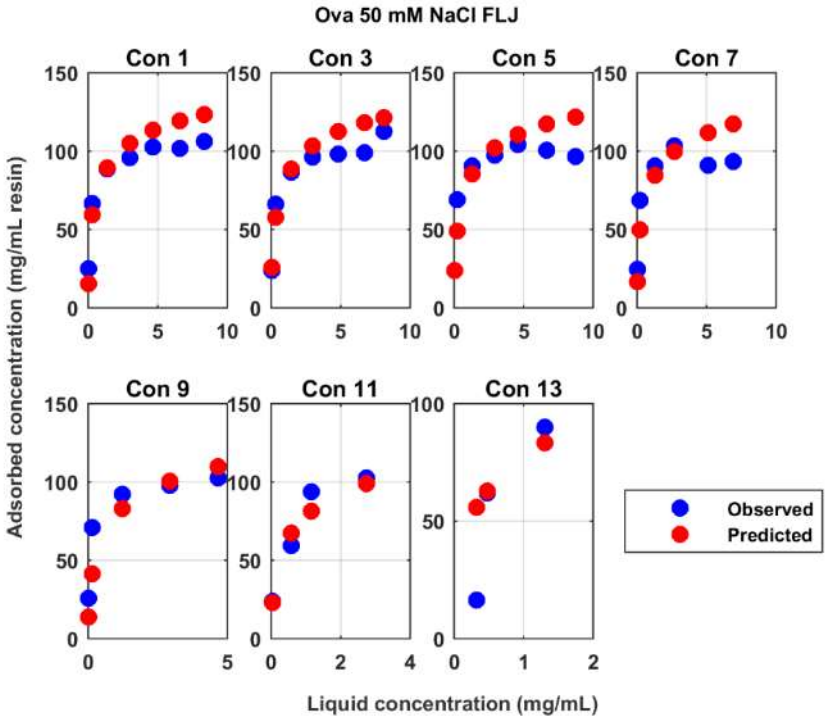

Figure 8-6 Freundlich-Langmuir-Jovanovic (FLJ) fit to binary Ova-Con binary isotherm showing observed and predicted values for Ova adsorbed concentration at 50 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of Con displayed as the title of each subplot.

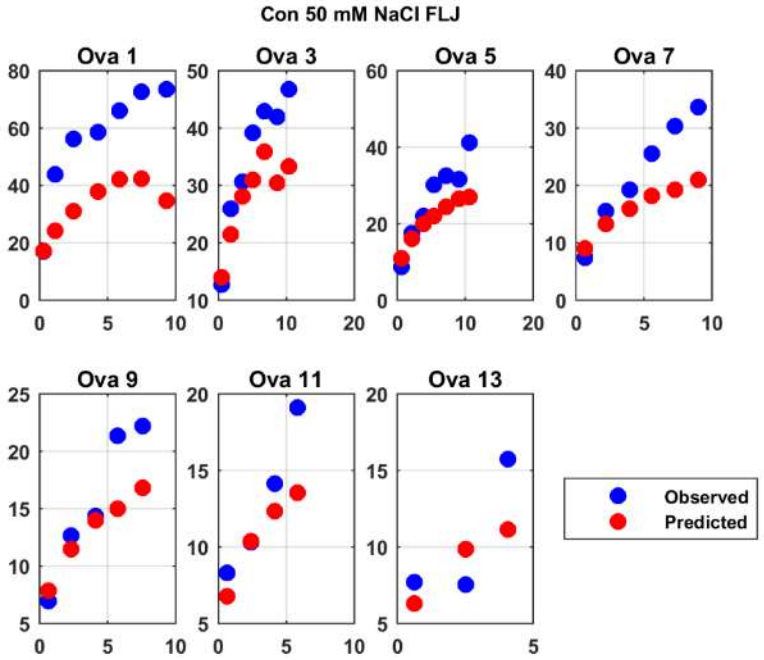

Figure 8-7 Freundlich-Langmuir-Jovanovic (FLJ) fit to binary Ova-Con binary isotherm showing observed and predicted values for Con adsorbed concentration at 50 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of Ova displayed as the title of each subplot.

8.2.1.3 BSA-Ova-Con ternary

The R^2 values for the ternary isotherm fit are low. It is reasonable to suggest that the best overall R^2 values for the three proteins is associated with the FLJ formalism. As such the quality of the isotherm fit will be discussed and is displayed in Figure 8-8 to Figure 8-10.

The FLJ isotherm systematically underestimates the adsorbed concentration of BSA when the starting concentration of competitor protein is low and the rectangular nature of the isotherm is not captured in these situations. As the BSA isotherm is supressed by either increasing starting concentration of competitor protein or increasing NaCl concentration agreement between the observed and modelled adsorbed concentration improves. Agreement between observed and modelled adsorbed concentration for Ova appears comparatively reasonable at 0 mM NaCl which is also reflected in the R^2 which is 0.6. At 25 mM NaCl the FLJ isotherm systematically underestimates the adsorbed concentration of Ova at lower starting concentrations of BSA, as the starting concentration of Con increases the FLJ isotherm systematically overestimates the Ova adsorbed concentration. However, the 25 mM isotherm observed data is also somewhat noisy. The 50 mM data set showing adsorbed Ova concentration is also rather noisy, the FLJ fit appears not to capture the displacement effect caused by BSA well at moderate starting concentration of Con, the fitted values nearly over lay in this region. Con adsorbed concertation estimates using the FLJ isotherm tend to show systematic overestimation at the lowest starting concentration of BSA, agreement otherwise is comparatively reasonable.

In summary the FLJ isotherm systematically mis-estimates much of the ternary and has particular difficulty capturing the trends in conditions of low competition where more rectangular behaviours are observed.

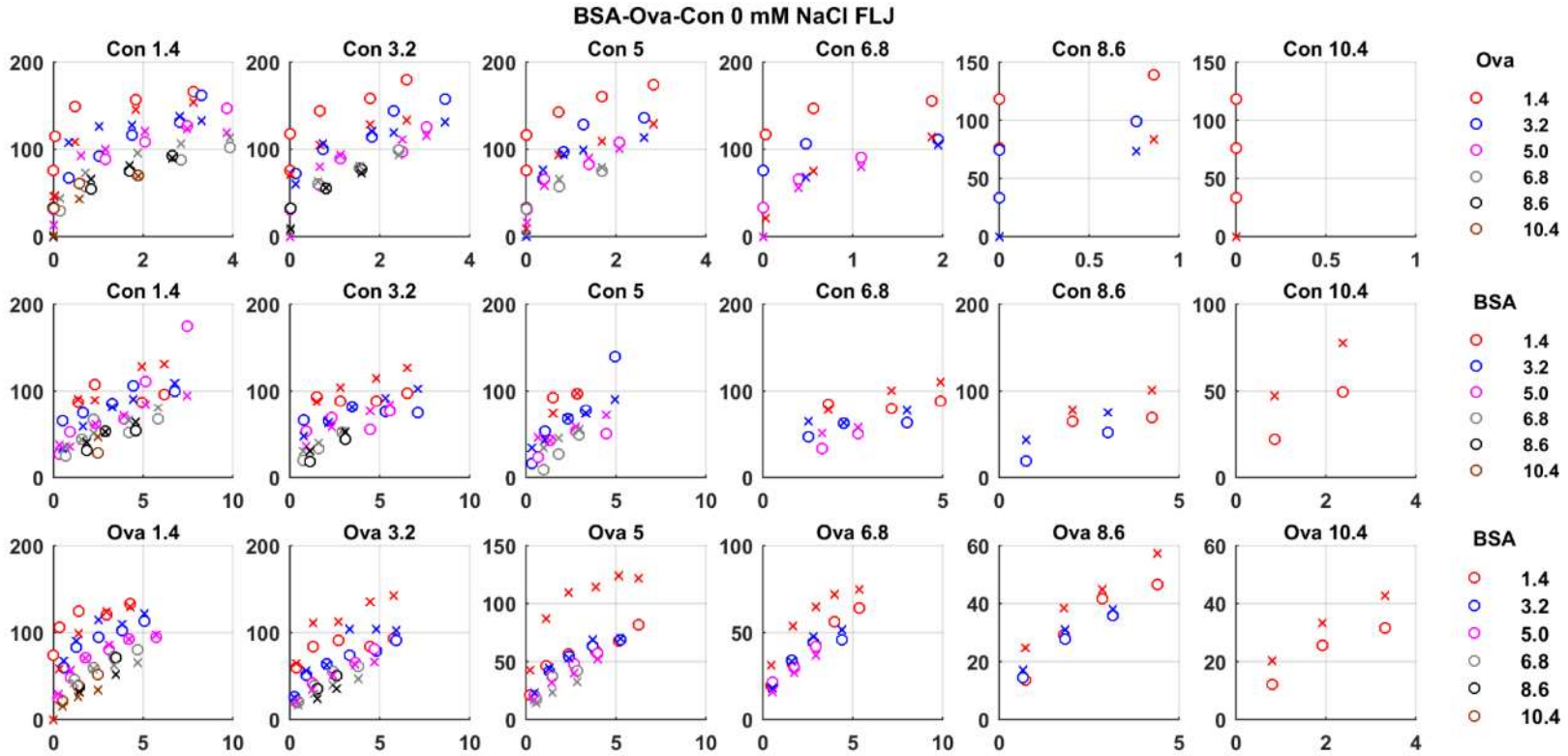

Figure 8-8 Ternary isotherm of BSA-Ova-Con isotherm with Freundlich-Langmuir-Jovanovic (FLJ) fits as a series of single component isotherms at different starting concentrations of competitor protein at pH 9 0 mM NaCl. The plot titles display the starting concentration of one of the competitors and the legends display the starting concentration of the second competing protein in mg/mL. All x axis display liquid concentrations in mg/mL and all y-axis represent adsorbed concentration in mg/mL of resin. Top row of 6 plots display BSA data, second row of 6 plots displays Ova data and the third row of 6 plots displays Con data. Circles represent experimental data and crosses represents FLJ fit.

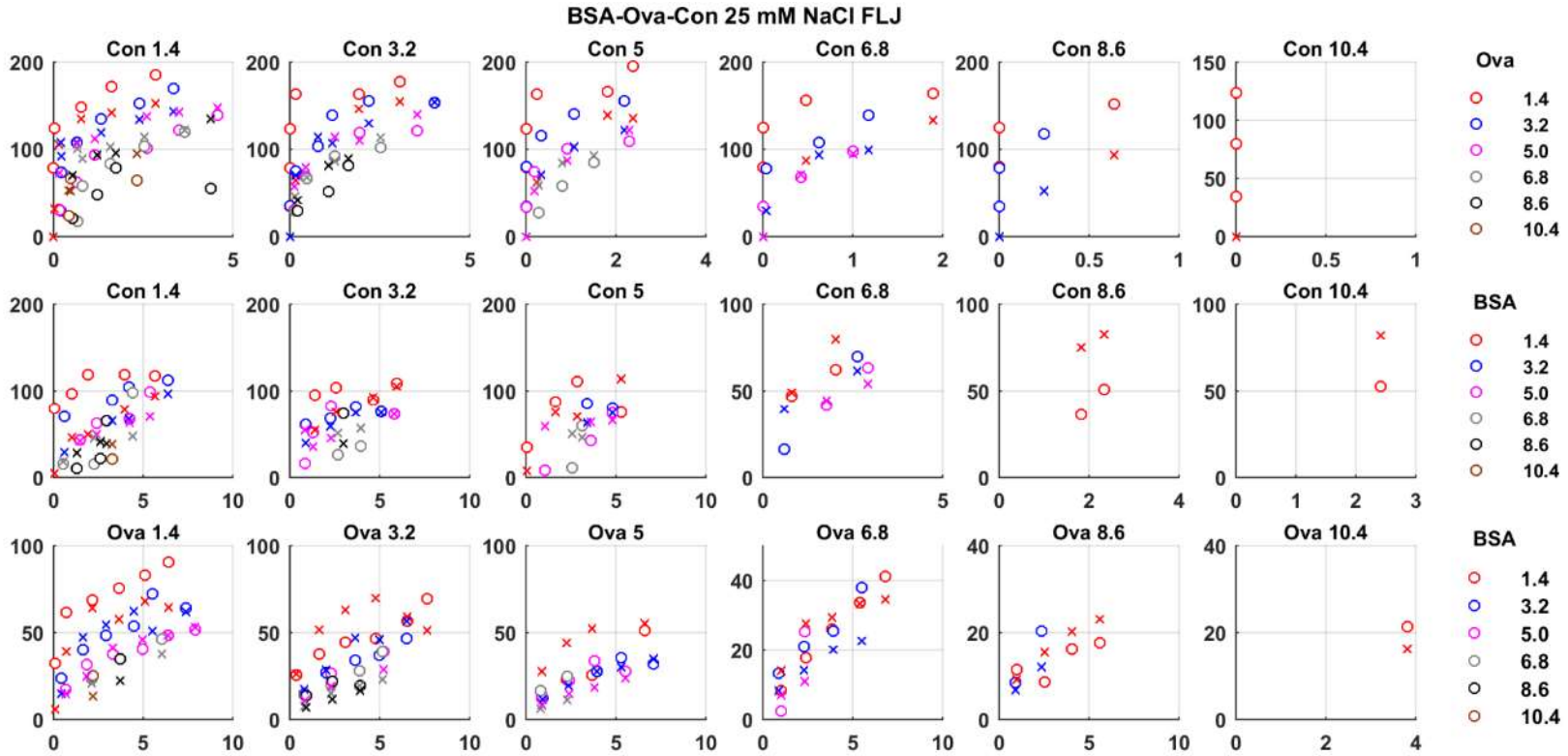

Figure 8-9 Ternary isotherm of BSA-Ova-Con isotherm with Freundlich-Langmuir-Jovanovic (FLJ) fits as a series of single component isotherms at different starting concentrations of competitor protein at pH 9 25 mM NaCl. The plot titles display the starting concentration of one of the competitors and the legends display the starting concentration of the second competing protein in mg/mL. All x axis display liquid concentrations in mg/mL and all y-axis represent adsorbed concentration in mg/mL of resin. Top row of 6 plots display BSA data, second row of 6 plots displays Ova data and the third row of 6 plots displays Con data. Circles represent experimental data and crosses represents FLJ fit.

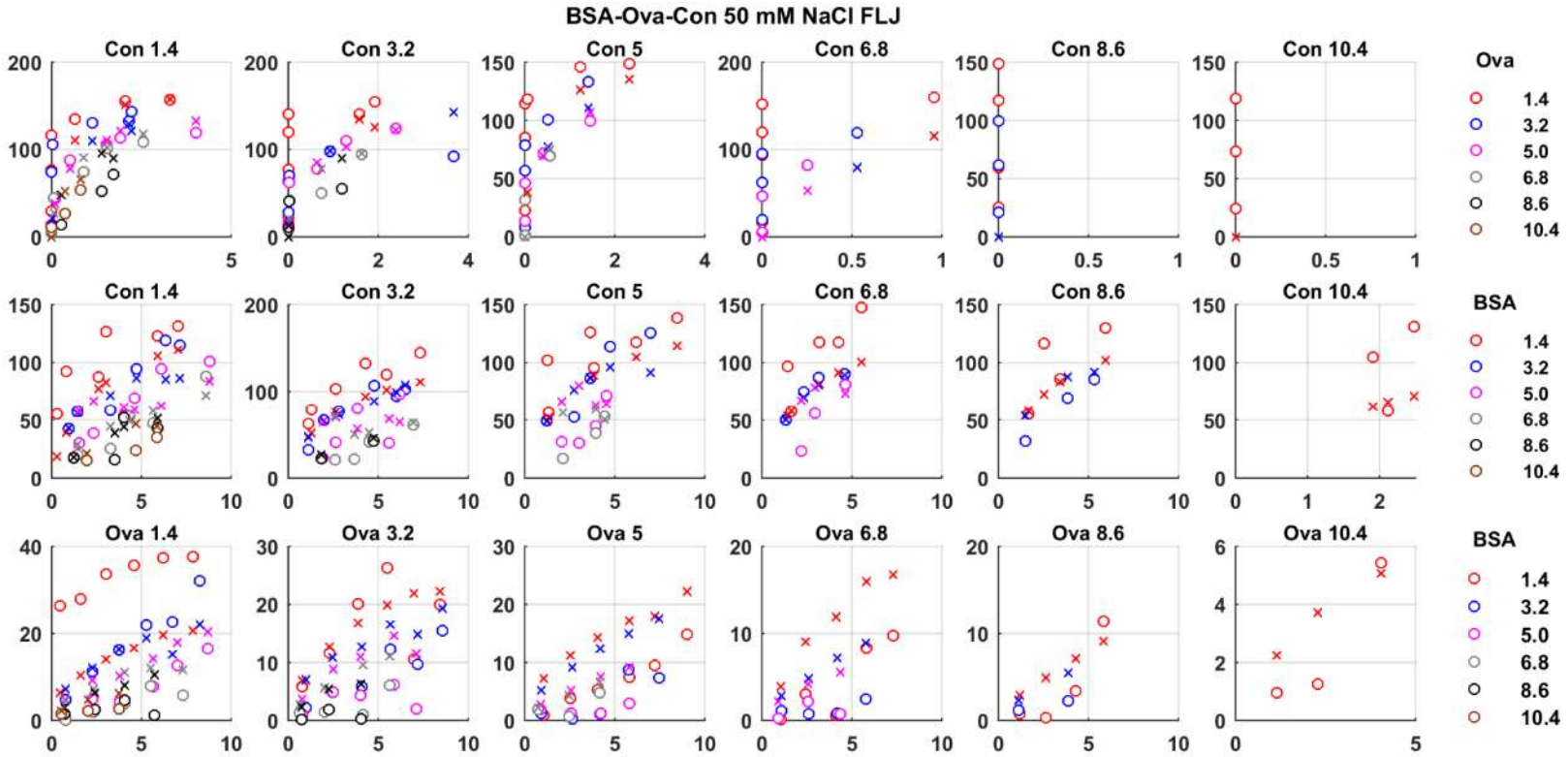

Figure 8-10 Ternary isotherm of BSA-Ova-Con isotherm with Freundlich-Langmuir-Jovanovic (FLJ) fits as a series of single component isotherms at different starting concentrations of competitor protein at pH 9 50 mM NaCl. The plot titles display the starting concentration of one of the competitors and the legends display the starting concentration of the second competing protein in mg/mL. All x axis display liquid concentrations in mg/mL and all y-axis represent adsorbed concentration in mg/mL of resin. Top row of 6 plots display BSA data, second row of 6 plots displays Ova data and the third row of 6 plots displays Con data. Circles represent experimental data and crosses represents FLJ fit.

8.2.2 Steric mass action isotherm fitting

Summary of fitting results is displayed in Table 8-1 with R^2 values displayed. Generally the fits using the SMA model are extremely poor with R^2 never better than 0.6. The isotherm best described by the SMA model is the BSA-Ova binary at pH 9, as such it will be discussed in greater depth below to ascertain where the formalism reflects the data well and where it is less successful.

8.2.2.1 BSA-Ova binary pH 9

SMA fits to the BSA-Ova binary at pH 9 in a range of NaCl levels is displayed Figure 8-11 to Figure 8-17. Figure 8-11 to Figure 8-14 display the observed and predicted adsorbed concentrations for BSA. The SMA model systematically mis-estimates the adsorbed BSA concentration at all but the highest level of 200 mM NaCl in Figure 8-14. At 0 mM NaCl in Figure 8-11 the SMA overestimates the adsorbed concentration at moderate starting concentrations of competitor, then at 50 (Figure 8-12) and 100 mM NaCl (Figure 8-13) it underestimates the adsorbed concentration at high starting concentrations of competitor.

Modelled adsorbed concentrations of Ova is displayed in Figure 8-15. Ova adsorbed concentration is severely mis-estimated at NaCl concentrations of 50 mM and above. At 50 mM NaCl (Figure 8-15) the SMA underestimates the adsorbed concentration at low starting concentrations of competitor, at 100 mM NaCl (Figure 8-16) the SMA underestimates the adsorbed concentration at low starting concentrations of BSA before crossing over and overestimating the adsorbed concentration. The overestimation also occurs at the highest NaCl concentration of 200 mM.

As also reflected in Table 8-1 the SMA model shows a poor fit for the BSA-Ova binary. There is severe systematic mis-estimation of adsorbed concentration for both BSA and Ova at nearly all NaCl concentrations studied. Both competitive behaviour of one protein displacing the other and the effect of NaCl on the isotherm is not captured reliably by the SMA model.

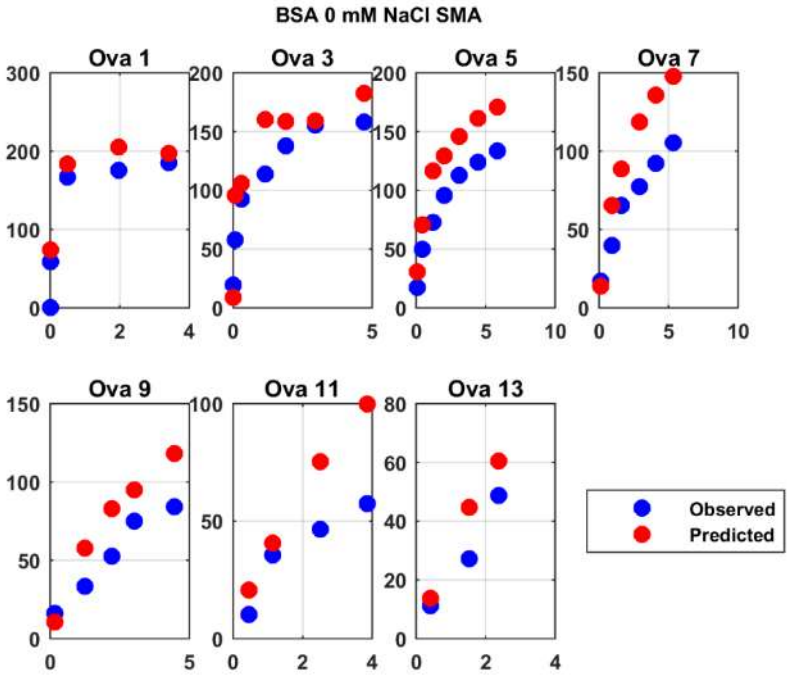

Figure 8-11 Steric mass action (SMA) fit to binary BSA-Ova binary isotherm showing observed and predicted values for BSA adsorbed concentration at 0 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of Ova displayed as the title of each subplot.

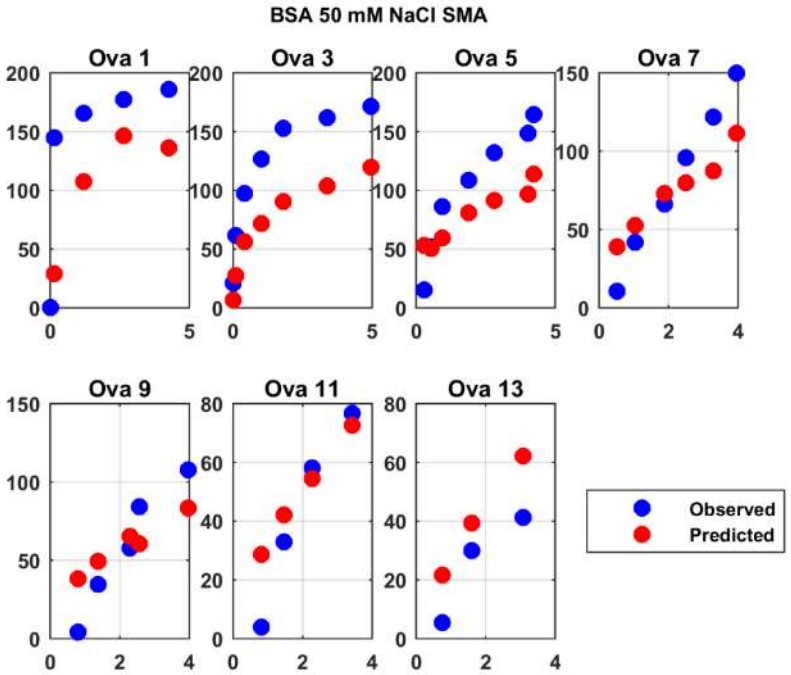

Figure 8-12 Steric mass action (SMA) fit to binary BSA-Ova binary isotherm showing observed and predicted values for BSA adsorbed concentration at 50 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of Ova displayed as the title of each subplot.

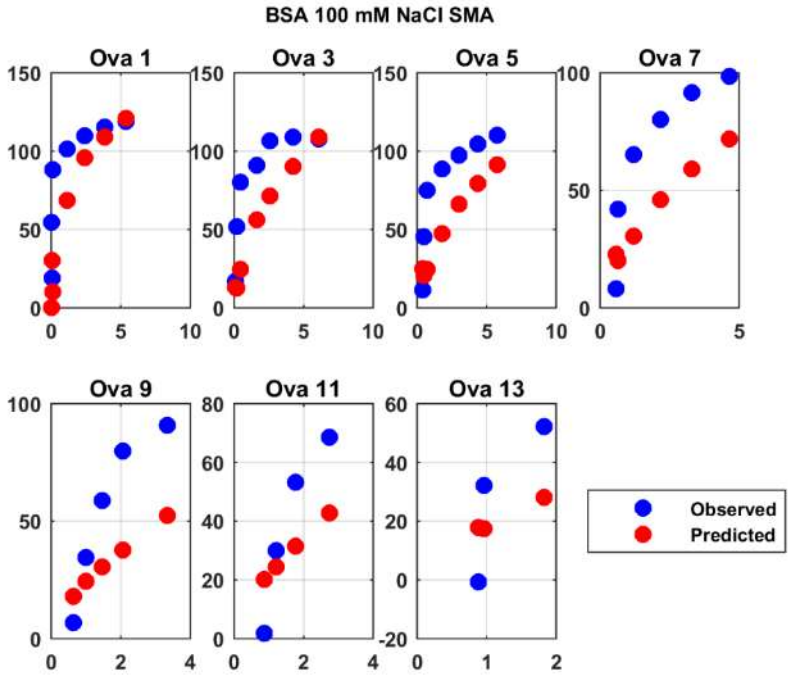

Figure 8-13 Steric mass action (SMA) fit to binary BSA-Ova binary isotherm showing observed and predicted values for BSA adsorbed concentration at 100 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of Ova displayed as the title of each subplot.

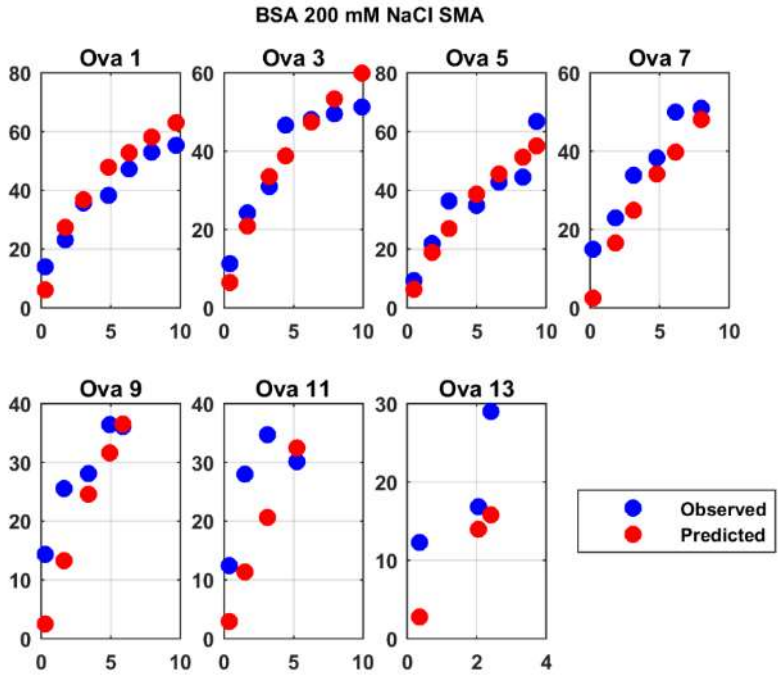

Figure 8-14 Steric mass action (SMA) fit to binary BSA-Ova binary isotherm showing observed and predicted values for BSA adsorbed concentration at 200 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of Ova displayed as the title of each subplot.

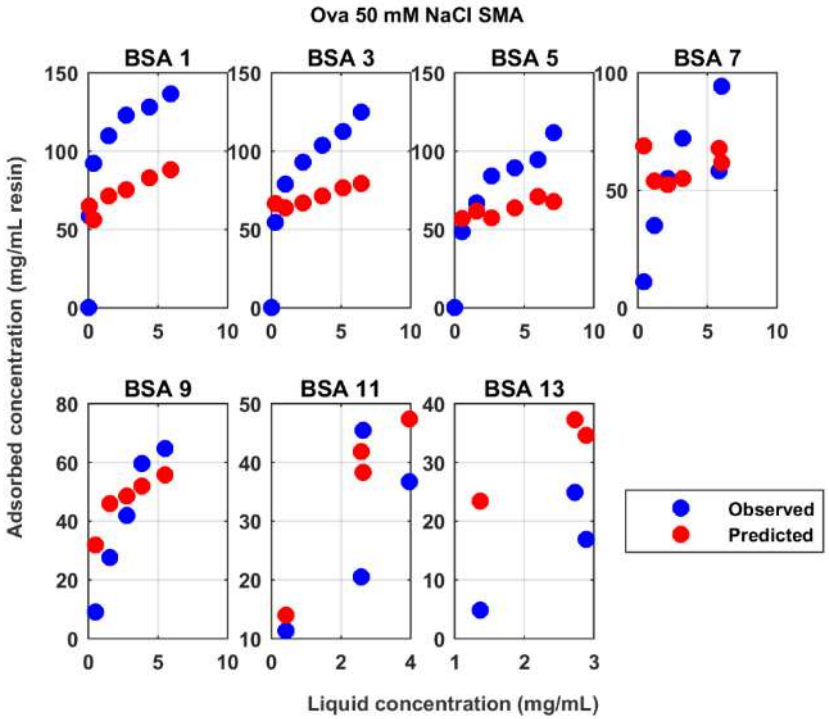

Figure 8-15 Steric mass action (SMA) fit to binary BSA-Ova binary isotherm showing observed and predicted values for Ova adsorbed concentration at 50 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of BSA displayed as the title of each subplot.

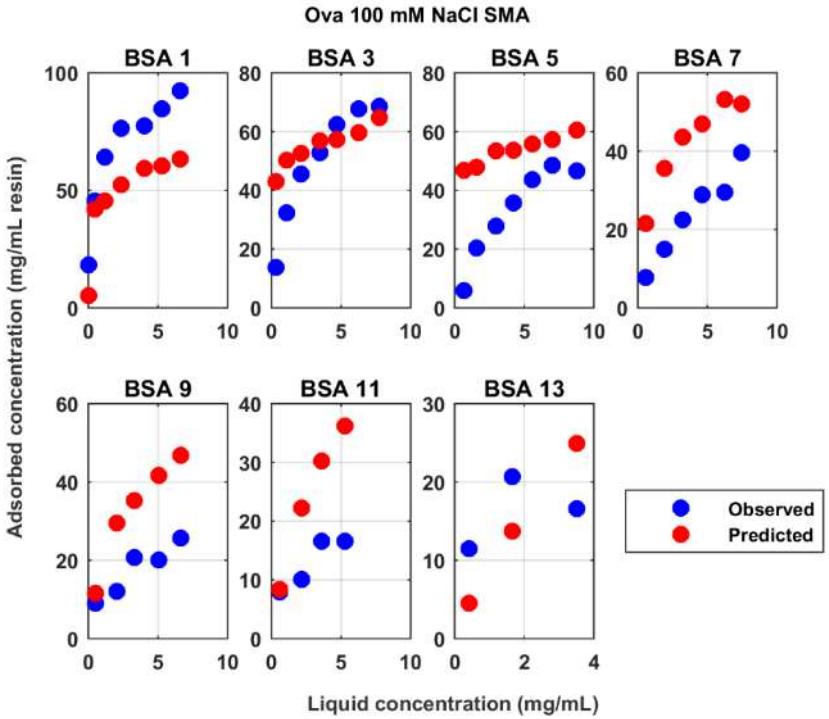

Figure 8-16 Steric mass action (SMA) fit to binary BSA-Ova binary isotherm showing observed and predicted values for Ova adsorbed concentration at 100 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of BSA displayed as the title of each subplot.

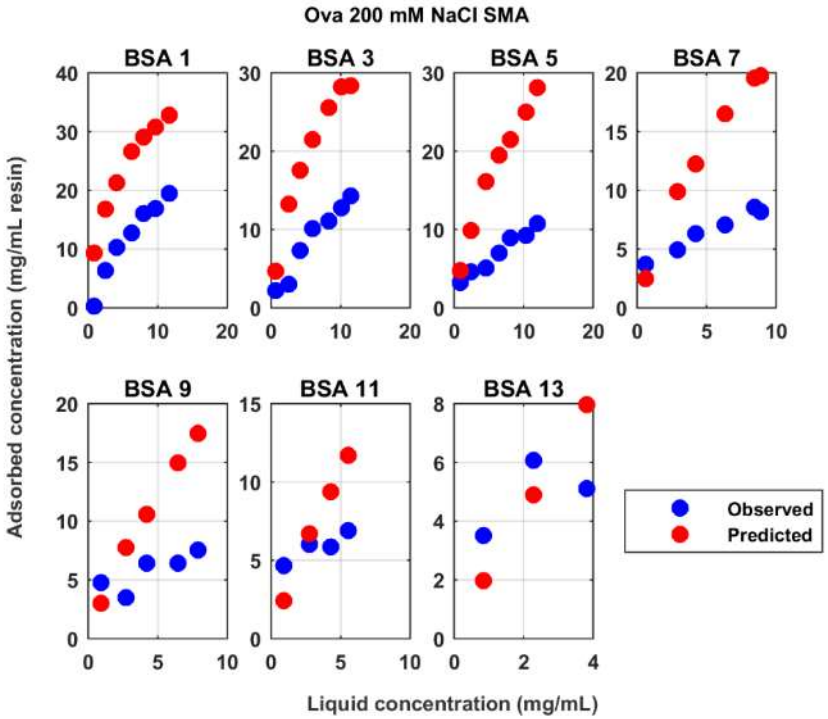

Figure 8-17 Steric mass action (SMA) fit to binary BSA-Ova binary isotherm showing observed and predicted values for Ova adsorbed concentration at 200 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of BSA displayed as the title of each subplot.

8.2.3 Padé isotherm fitting

A summary of isotherm fits is displayed in Table 8-1 along with R^2 values for the Padé isotherm fit. The overall R^2 values are generally quite poor for the Padé description R^2 only exceeding 0.6 for BSA-Ova binary isotherm as well as in the case of Ova in the Ova-Con binary. Closer inspection of the R^2 values shows that in the case of the BSA-Ova binary, the R^2 value is only 0.01 for Ova at 200 mM NaCl. To ascertain how well the Padé description describes the trends seen in the BSA-Ova the fits will be presented and discussed below.

Figure 8-18 to Figure 8-25 display the Padé isotherm fit to BSA-Ova binary at multiple NaCl levels. The figures corroborate the R^2 values observed in Table 8-1 suggesting the fits are in good agreement with the data. There is very little systematic mis-estimation throughout the isotherm for estimates of adsorbed concentration for both BSA and Ova as shown in Figure 8-18, Figure 8-19 and Figure 8-20 with the exception of Ova at 200 mM NaCl (Figure 8-21) for which the adsorbed concentration is systematically underestimated. This is corroborated by the R^2 which was poor with a value of 0.01 in this instance.

The Padé surface fit to the experimental data in 3 dimensions with equilibrium liquid concentration of BSA and Ova plotted against the adsorbed concentration of either BSA or Ova is displayed in Figure 8-22 to Figure 8-25. In Figure 8-22 which displays the Padé surface at 0 mM NaCl the BSA half of the binary isotherm shows implausible behaviour. When the liquid concentration of BSA is extrapolated beyond the experimental data anti-Langmuirian behaviour is observed as discussed in section 2.2 Fundamentals solute-sorbent interaction at equilibrium and the effect mass transfer resistance, this behaviour is highly unusual in solid liquid chromatography, particularly in the case of a protein solute. Behaviour such as this would equate to multilayer binding of BSA on the sorbent with the protein-protein interactions having increasing interaction strength with increasing coverage. Additionally, as this behaviour is observed in a region of extrapolation with no direct experimental observations to corroborate it. This suggests that this behaviour is an artefact of the

fit rather than a reflection of the data. Similar unusual behaviour in modelled adsorbed BSA concentration is also observed in Figure 8-23. Additionally, Figure 8-24 which shows experimental data and the Padé surface at 100 mM NaCl exhibits non-monotonic behaviour when looking at modelled adsorbed Ova concentration. This artefact also occurs in a region of extrapolation and is not corroborated by any experimental data. A Padé surface for an NaCl level for which there is no experimental data is shown in Figure 8-25, the surface continues to display the anti-Langmuirian behaviour exhibited at 50 mM in Figure 8-23.

Although the Padé isotherm formalism has R^2 values associated with it which suggesting a good fit to the data, inspection of the surfaces shows the Padé is unsuitable. When extrapolating only a small amount outside the observed equilibrium liquid protein concentration the formalism exhibits highly unlikely behaviour. Interpolation of surfaces at NaCl levels with no experimental data also exhibit unlikely behaviour. If *in silico* separations were performed using this isotherm description highly unusual peak shapes and behaviours would be observed making this isotherm description unsuitable despite the high R^2 values observed.

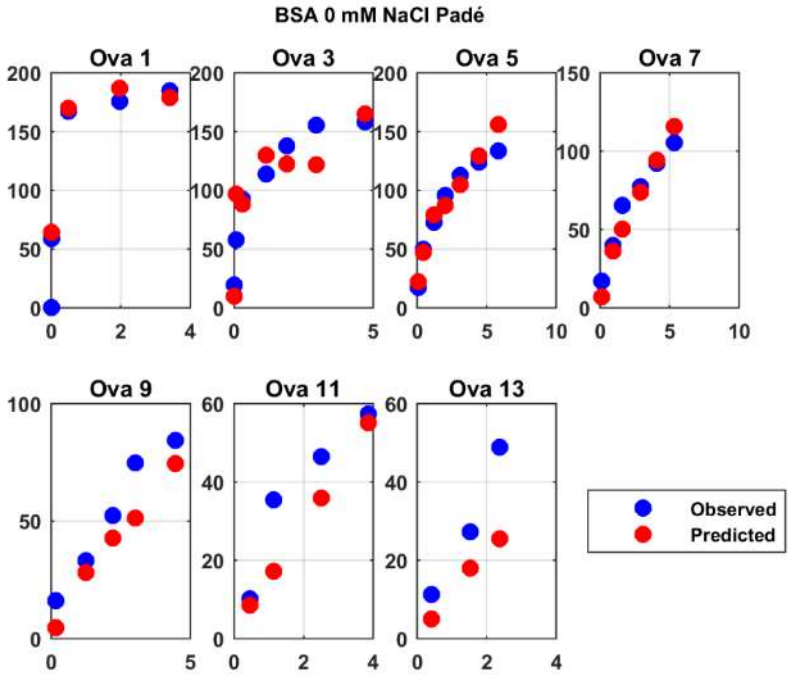

Figure 8-18 Padé fit to binary BSA-Ova binary isotherm showing observed and predicted values for BSA adsorbed concentration at 0 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of Ova displayed as the title of each subplot.

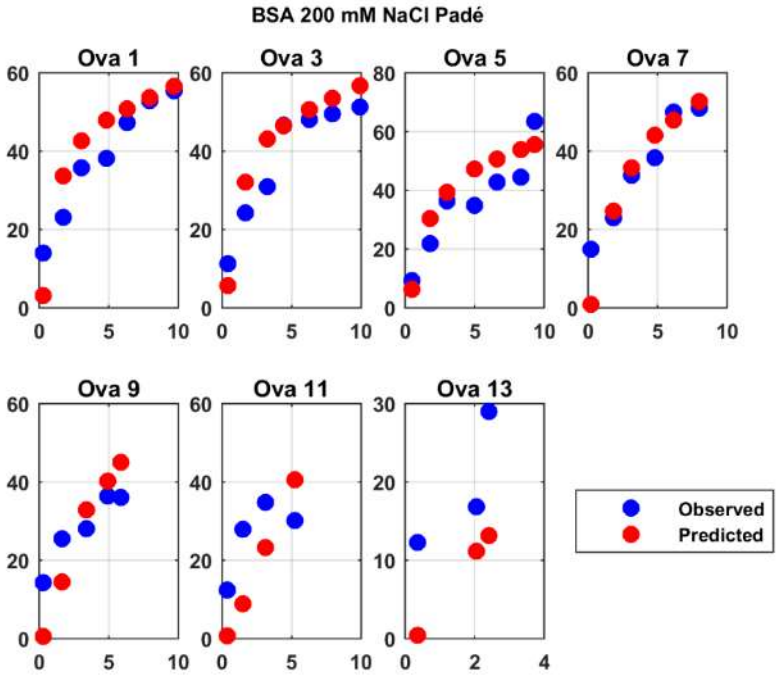

Figure 8-19 Padé fit to binary BSA-Ova binary isotherm showing observed and predicted values for BSA adsorbed concentration at 200 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of Ova displayed as the title of each subplot.

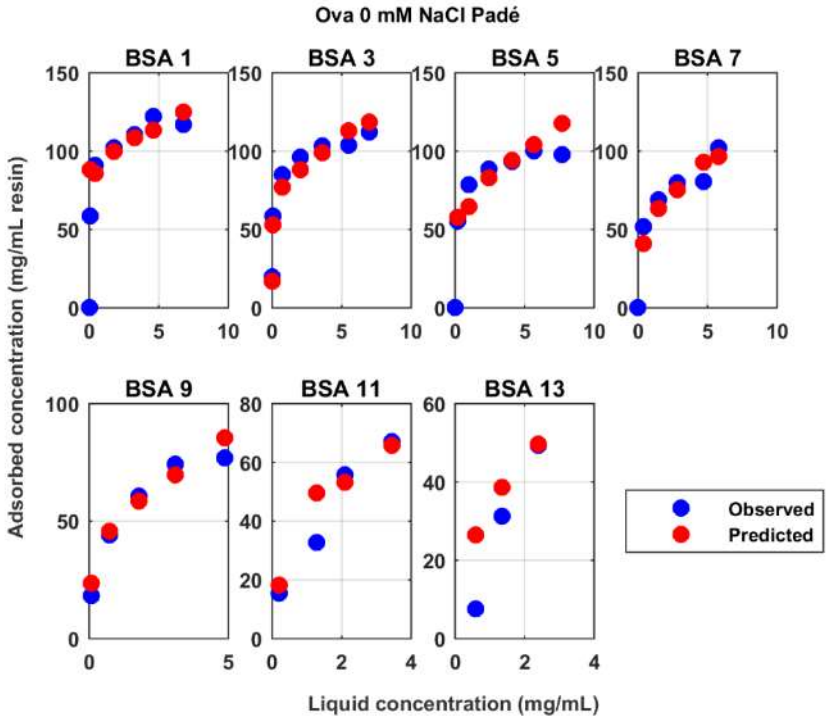

Figure 8-20 Padé fit to binary BSA-Ova binary isotherm showing observed and predicted values for Ova adsorbed concentration at 0 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of BSA displayed as the title of each subplot.

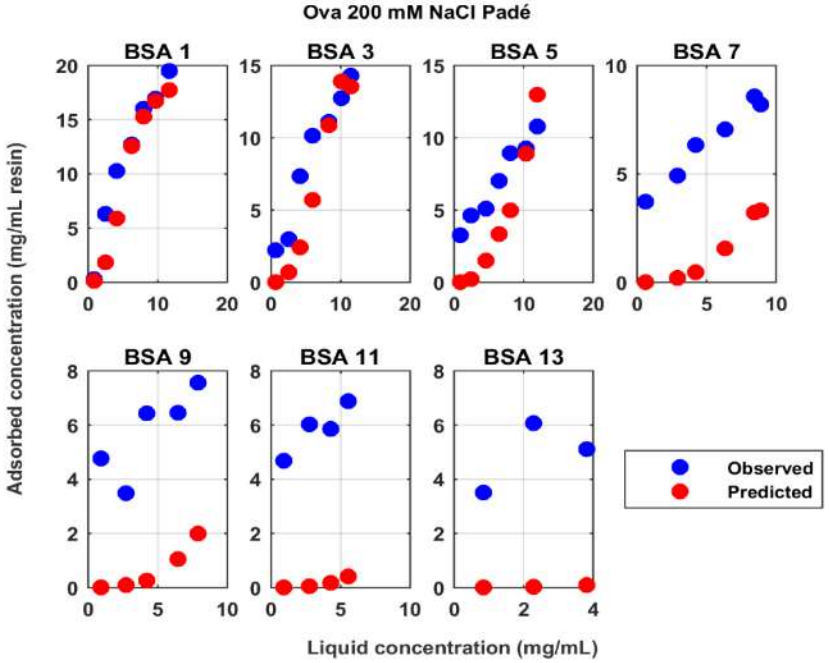

Figure 8-21 Padé fit to binary BSA-Ova binary isotherm showing observed and predicted values for Ova adsorbed concentration at 200 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of BSA displayed as the title of each subplot.

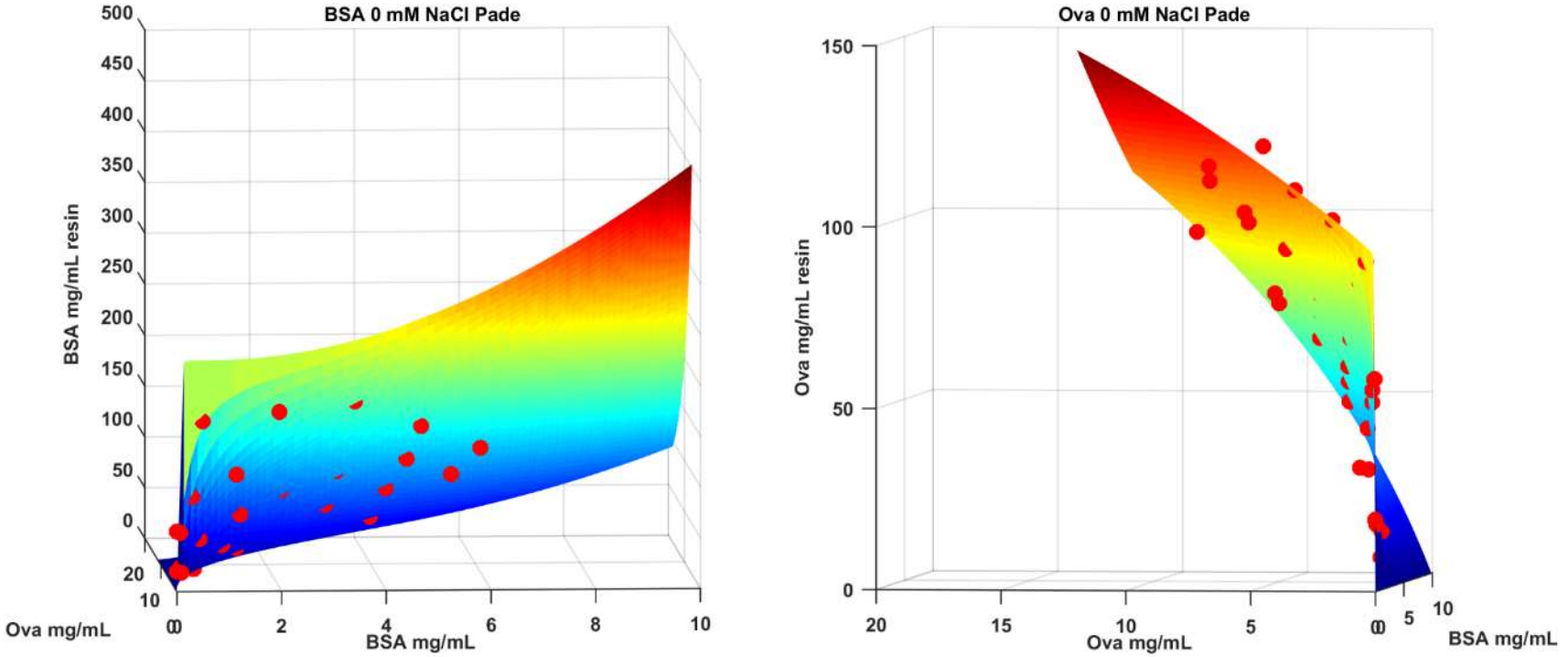

Figure 8-22 BSA-Ova binary isotherm displaying the Padé surface fit at 0 mM NaCl. Left hand panel displays BSA adsorbed concentration and right hand panel adsorbed Ova. Experimental data has been ploted as red points.

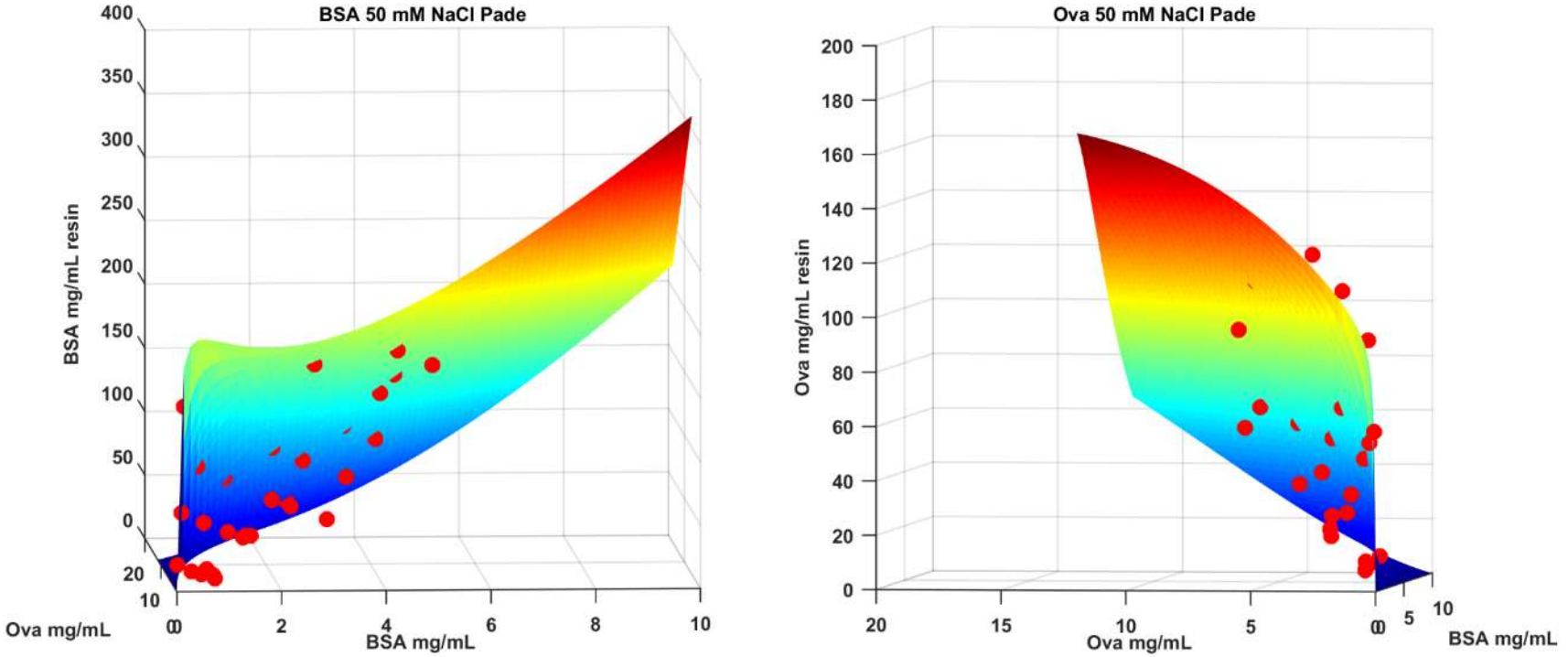

Figure 8-23 BSA-Ova binary isotherm displaying the Padé surface fit at 50 mM NaCl. Left hand panel displays BSA adsorbed concentration and right hand panel adsorbed Ova. Experimental data has been ploted as red points.

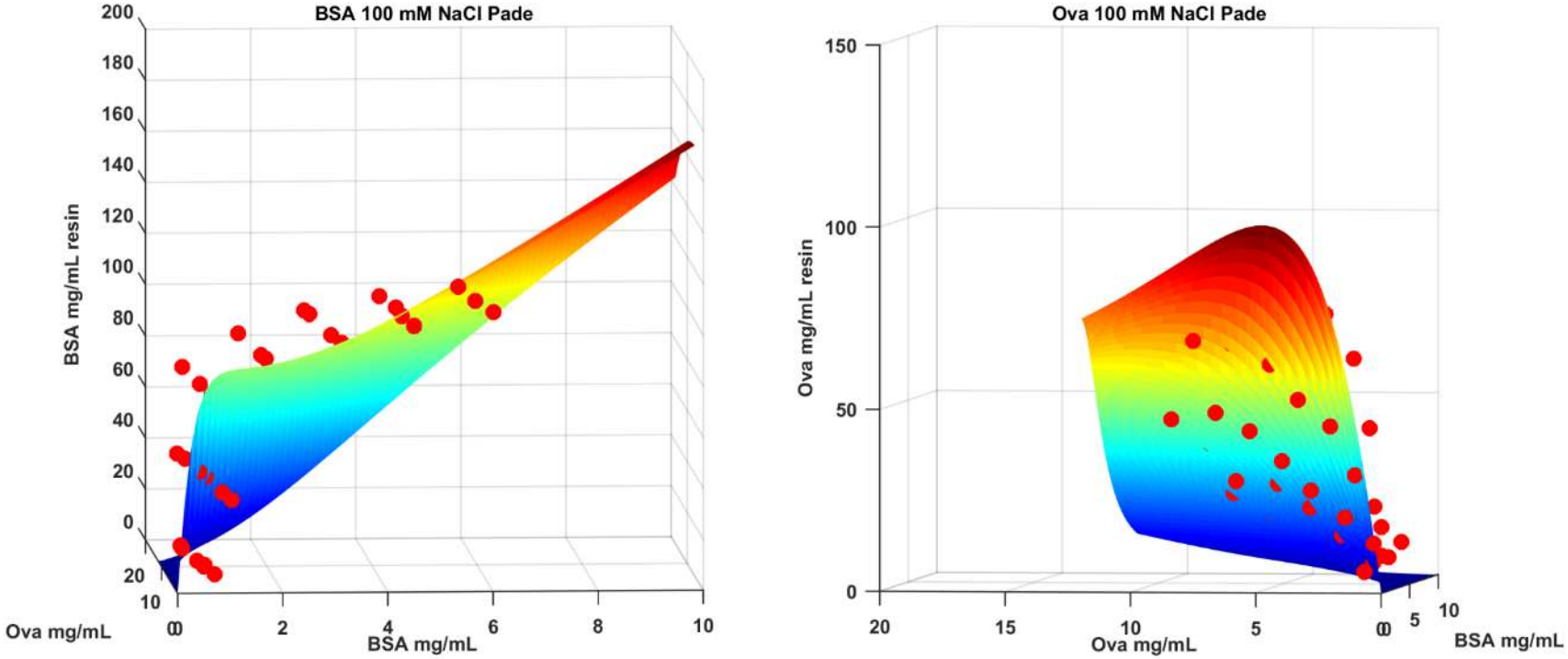

Figure 8-24 BSA-Ova binary isotherm displaying the Padé surface fit at 100 mM NaCl. Left hand panel displays BSA adsorbed concentration and right hand panel adsorbed Ova. Experimental data has been ploted as red points.

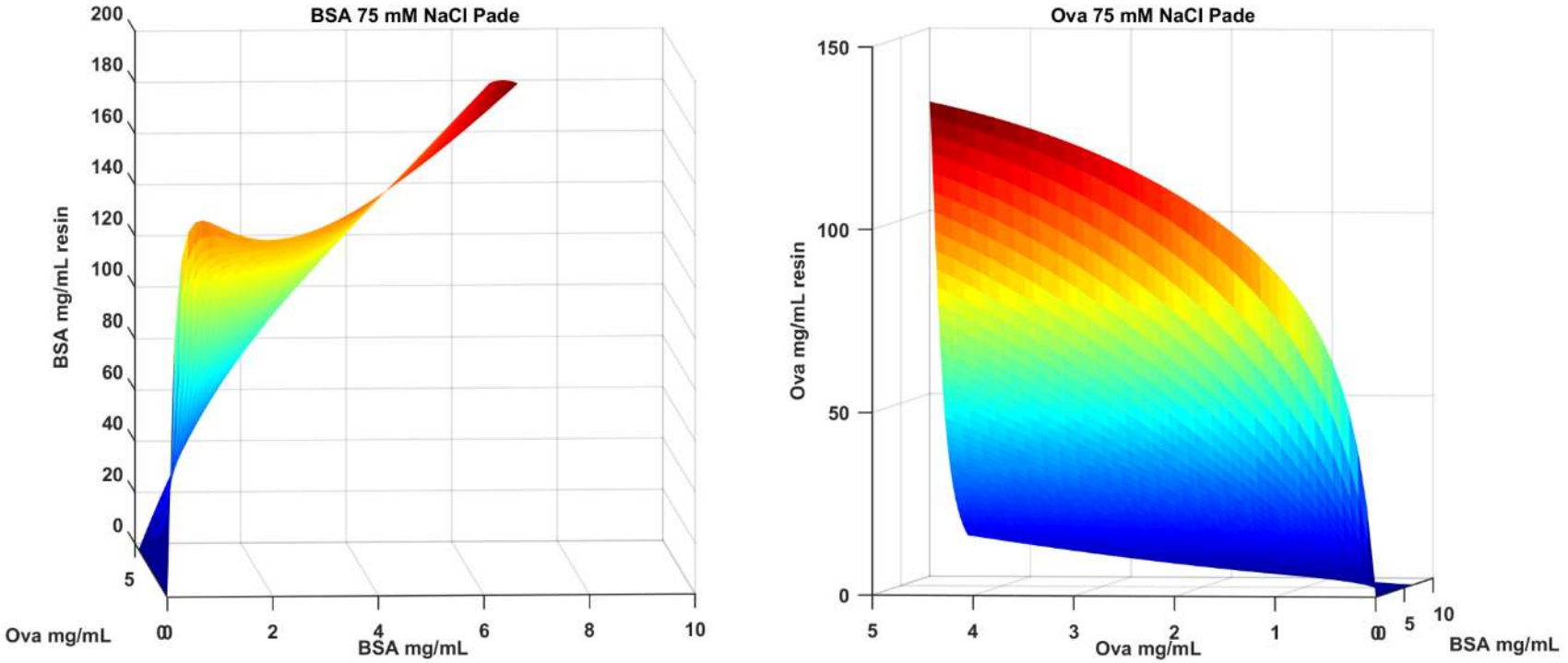

Figure 8-25 BSA-Ova binary isotherm displaying the Padé surface fit at 75 mM NaCl. Left hand panel displays BSA adsorbed concentration and right hand panel adsorbed Ova. There is no experimental data to display at this NaCl level.

8.3 Discussion and summary

Fitting the multicomponent isotherms to isotherm formalisms yielded little success. The FLJ, SMA and Padé isotherms all displayed significant systematic mis-estimation in adsorbed protein concentration. The FLJ isotherm did capture trends observed in the Ova-Con isotherm with comparatively little systematic mis-estimation. The Ova-Con isotherm exhibits some features which potentially mean it is easier to fit to a formalism. Figure 4-11 showed that the Ova isotherm exhibits less severe rectangular behaviour than both BSA and Con at low NaCl concentrations. It is well know that capturing highly rectangular behaviour using isotherm formalisms is challenging. Additionally the multicomponent behaviour of the BSA-Con isotherm is comparatively simple with Ova being minimally displaced by Con meaning competitive behaviours potentially less challenging to capture. The BSA-Ova isotherm shows comparatively complex displacement behaviour for both proteins. Additionally, the Con-BSA binary shows highly rectangular behaviour for both proteins. In conjunction these factors may mean that the Ova-Con isotherm is the most facile to capture. Capturing the more complex behaviours seen in other multicomponent isotherms could potentially be achieved with the use of a more complex formalism which uses more parameters such as the Padé.

The additional parameters in the Padé isotherm description were found to replicate the observed experimental data quite well for the BSA-Ova binary. However, on inspection of the Padé surface it was also associated with isotherm behaviours which are unlikely when conservative extrapolation was explored. Moreover, interpolation of adsorbed Padé isotherm surfaces with NaCl levels also showed unfeasible behaviour. The potential power of the Padé description in being able to replicate complex observed behaviours using many more parameters also means the formalism can explore unlikely shapes on extrapolation. The more traditional isotherm formalisms explored such as the SMA and FLJ descriptions have fewer parameters to explore complex behaviours but they are also constrained

meaning the unlikely behaviours exhibited by the Padé are not accessible on extrapolation outside of the experimental space.

As all the isotherm formalisms explored in this chapter showed significant systematic mis-estimation *in silico* modelling of multicomponent separations is not feasible. Work not discussed here attempted to model BSA-Ova separation using these fits did show some success but were unable to fully replicate behaviour on columns. Fitting complex multicomponent behaviours to isotherm formalisms is known to be a challenging task (Abdehagh, Tezel, & Thibault, 2016; Liang et al., 2012; Reck, Pabst, Hunter, Wang, & Carta, 2015; Wu & Lin, 2009; Xu & Lenhoff, 2009). The challenge is likely made even more difficult as the proteins explored here first exhibit complex competitive behaviours with BSA and Ova displacing one another with similar effectiveness. Additionally, these proteins are larger than the typical model proteins such as cytochrome-c and lysozyme meaning they may potentially exhibit more complex behaviours. Finally, the high density and variety of protein loads along with the exploration of multiple NaCl levels explores the competitive and suppressive nature of competition in adsorption in increased detail. Whilst this added level of detail allows closer inspection of the behaviours overserved it also makes capturing those same details more challenging.

The ternary data which was used in order to fit isotherm formalisms here is extremely rare in the literature, we are not aware of any such data for such closely retained compounds at such a high level of granularity. The multi component isotherm data is useful in selecting isotherm formalisms that fit well, and de-emphasise formalisms which fit poorly. Unfortunately no isotherm formalisms were found which fit the ternary data well. The results did show that both the SMA and Padé were unable to capture the trends, and were therefore not useful for this system.

Chapter 9 Optimisation of BSA-Ova binary separation

9.1 Aims and objectives of chapter

The aim of this chapter is to present the optimisation of BSA-Ova separation on Capto^{TM} Q at pH 9. Ideally the binary isotherms generated would have been used to model and optimise the separation *in silico*. However, none of the isotherm descriptions explored gave acceptable fits and *in silico* models did not perform acceptably. However, the information gained from the multicomponent isotherm can still be used to aid decision making in the optimisation of separation. Additionally, the high throughput analysis method involving UV spectra and PLS models explored in Chapter 5 is used to analyse elution fractions and expedite the optimisation process. The specific aims of the chapter are to:

- Demonstrate the use of PLS models in the analysis of elution fractions.
- Demonstrate the heuristic optimisation of BSA-Ova mixtures at pH 9.

9.2 Quantification of BSA-Ova column run fractions

BSA-Ova binary column runs were quantified by fractionating the eluate and analysing it on 96 well UV plates before processing it using the PLS models discussed Chapter 5 and used in Chapter 6.

Deconvolution of a BSA-Ova binary column runs is shown in Figure 9-1. The top panel displays the raw quantification of fractions using UV spectra in conjunction with PLS models. There are clearly artefacts in the quantification method with negative concentrations of BSA as well as the appearance of a BSA pre-peak immediately before Ova starts to elute. Negative concentrations are not physically possible and so have been set to 0 mg/mL in the bottom panel of Figure 9-1. All of the early eluting BSA fractions were pooled and analysed on HPLC along with BSA and Ova standards. The results of those HPLC runs are overlaid and shown and displayed in Figure 9-2. Figure 9-2 shows that there is no BSA present in the early eluting BSA

fraction in the top panel of Figure 9-1. As such, not only have the negative concentrations been ignored in the bottom panel of Figure 9-1 but concentrations of BSA pre-peak have been set to 0 mg/mL. As the HPLC analysis showed that there is only Ova present in the BSA pre-peak those fractions have been quantified using 280 nm data and the Ova standard curve.

After artefact removal, mass balance closure of BSA and Ova is 91% and 94% respectively suggesting the combination of PLS quantification with the checking of artefact presence with HPLC is fit for purpose. All binary runs discussed in this chapter displayed BSA pre-peaks which were shown to be artefacts after HPLC analysis as well as negative concentrations in solution. As such this methodology was applied throughout.

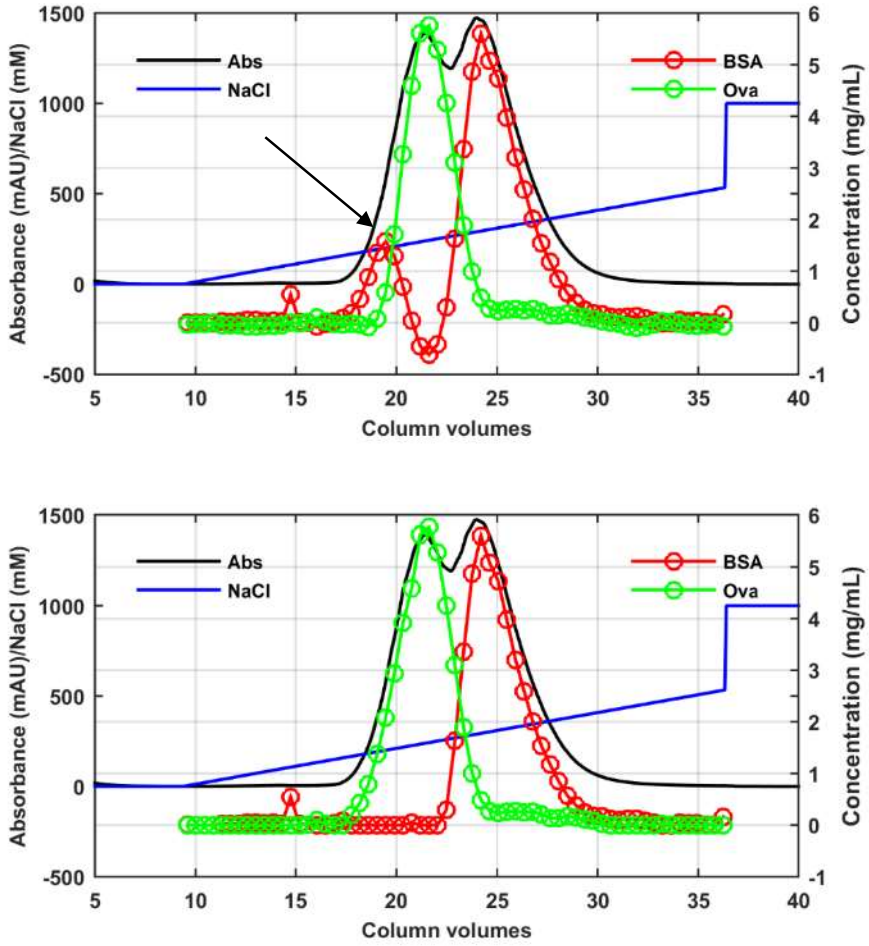

Figure 9-1 Results from deconvolution of BSA-Ova binary runs using UV spectra and PLS for quantification of fractions. Top panel shows the original results with an arrow pointing out the positive concentration artefact and the bottom panel after the artefacts have been removed.

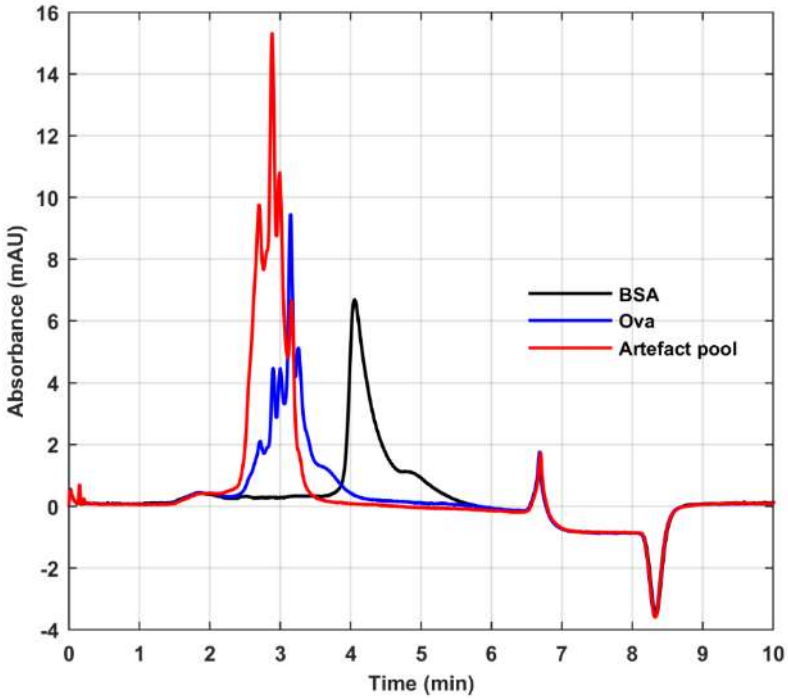

Figure 9-2 HPLC analysis of BSA and Ova standards and a sample which is the pool of the BSA artefact displayed in the top panel of (pooled 18-20 CV) Figure 9-1.

9.3 Optimisation of BSA-Ova separation

The aim of the optimisation is to maximise purity, yield and mass of Ova whilst maintaining purity ≥ 90% for Ova from a BSA-Ova binary load at a ratio of 1:1, NaCl gradient slope and load will be varied. An initial run was set up using a gradient slope of 40 mM/CV and a total load of 80 mg/mL resin. The initial load was chosen based on the batch adsorption studies which found that at 0 mM NaCl a total load of 120 mg/mL resulted in isotherms for both BSA and Ova being in the linear range where there was nearly no solute left in the liquid phase at equilibrium meaning very strong binding for both proteins. As a rule of thumb, dynamic binding capacity (DBC) at reasonable flow rates is ≈ 80% static binding capacity (Staby & Jensen, 2001) meaning a total load of 80 mg/mL resin is within sensible range. The initial NaCl did vary between 0 and 25 mM across the runs, it is assumed the low sensitivity of BSA and Ova in the isotherms to NaCl at low concentrations means these differences will minimally affect the outcome of the optimisation.

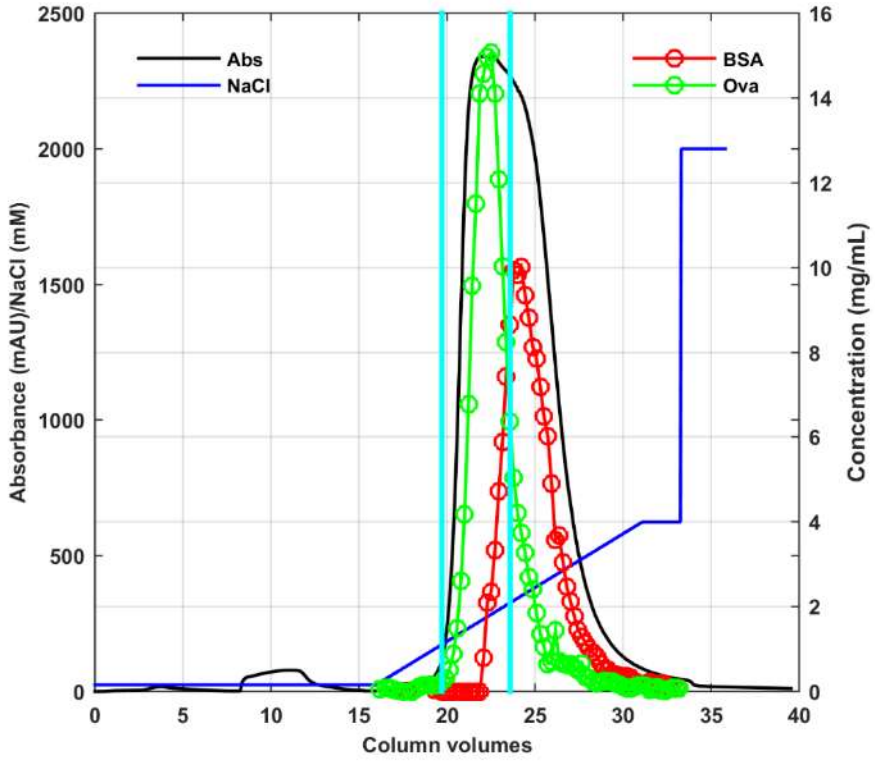

Figure 9-3 Deconvolution of initial BSA-Ova binary run 1. Cyan lines represent pooling choice. Run conditions were: Load 80 mg/mL resin and gradient of 40 mM/CV.

Results of the initial run after artefacts have been removed are displayed in Figure 9-3. There was a small amount of break through on loading but analysis of the load and wash fractions shows it only represents 4% of the total load. The purity and yield of Ova calculated after pooling was 81% and 74%, mass balance closure was relatively poor for BSA at 78% as displayed in Figure 9-4 which displays a summary of all the conditions and results. More accurate quantification to allow mass balance closure closer to 100% would have likely given even lower Ova purity results. The summary of all run conditions and results in Figure 9-4 suggests the slope of 40 mM/CV appears to be too steep to resolve BSA and Ova adequately to achieve > 90% purity of Ova. Even a decreased total load of 10 mg/mL resin in run 2 gives improved purity but still less than 90% with a slope of 40 mM/CV. Decreasing the slope of the separation at the same load of 10 mg/mL improves the purity which now exceeds 90% in run 3. Increased load in run 4 improved P*Y*M whilst maintaining purity > 90%, further increases in load decrease the purity as the peak width increases in runs 5 and 6.

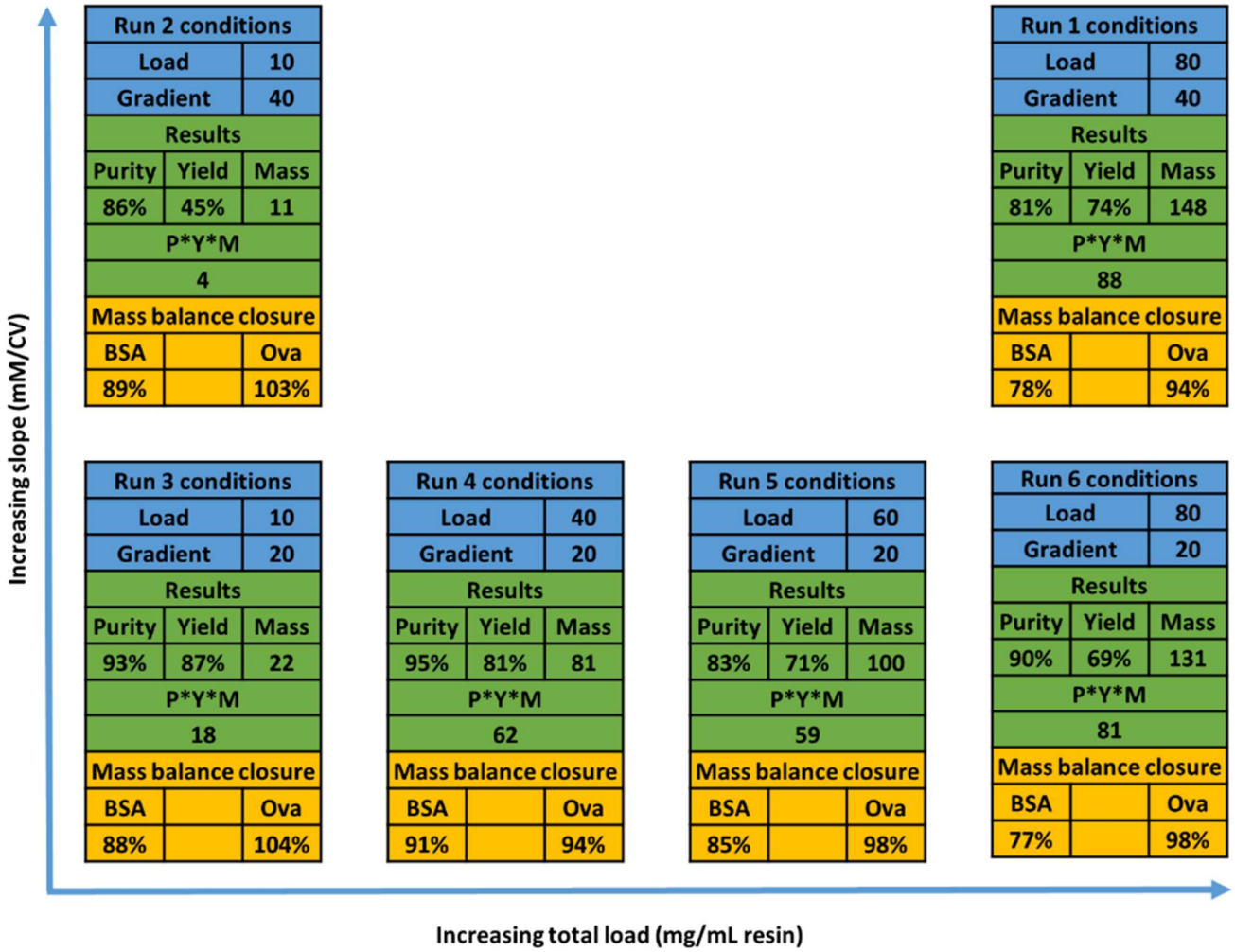

*Figure 9-4 Schematic summarising the conditions and results from BSA-Ova optimisation varying load and gradient slope. Load is in mg/mL resin (resin volume is 4.7 mL) gradient in mM/CV and mass in mg. P*Y*M is the purity multiplied by the yield and mass.*

9.4 Summary and discussion

The suggested optimal condition varying only slope and load for BSA-Ova separation in order to achieve Ova purity ≥ 90% purity for Ova and maximising P*Y*M using a 1:1 feed ratio BSA:Ova is a total load of 40 mg/mL resin and a slope of 20 mM/CV. Increasing load has the effect of increasing peak width and decreasing purity, increasing slope has the same effect.

The application of UV spectra in conjunction with PLS models to quantify the elution fractions from column runs was applied with relative success. There was the appearance of some negative concentrations for certain fractions which can justifiably be ignored. There was additionally the appearance of a BSA pre-peak before the elution of Ova which has been shown to be an artefact via identification via HPLC analysis. Although not discussed in detail the appearance of artefacts when using UV spectra in conjunction with PLS models does appear in the literature (Hansen et al., 2013). Mass balance closure of the fractions from the column runs is in excess of 85% for all but two of the highest loading column runs explored here. Mass balance closures of 86% have also been reported in the literature when using UV spectra and PLS model to quantify fractions (Hansen et al., 2011). When all the results are in context the poorer mass balance closure of these high loading runs is not expected to have altered the conclusions drawn.

Fitting of multicomponent isotherm data to formalisms did not yield acceptable fits to allow *in silico* modelling of the BSA-Ova separation. Work not presented here could not replicate observations made on columns, Failing that the batch adsorption data was used to aid choices in loading of the BSA-Ova separation and BSA-Ova separation was optimised heuristically.

Chapter 10 Summary and conclusions

10.1 Overall discussion and conclusions

Multicomponent chromatography adsorption isotherms have not been extensively studied historically. Frontal analysis methods are not well suited to their study, leaving the batch adsorption method as the best suited for their elucidation. Recent advances in high-throughput (HT) platforms have yielded slurry plates which represent a scaled down parallelised version of the batch adsorption experiment. This format used in conjunction with automated liquid handling systems (ALHS) allows for further benefits to be explored such as improved accuracy, repeatability and the possibility of automation.

The reasons why multicomponent isotherm studies of proteins have not been extensively studied include the requirement for many experimental points to be carried out as well as the propensity for error in such studies (Xu & Lenhoff, 2009). Whilst slurry plates increase the throughput of points which can be generated for such studies they do not alleviate the analytical bottleneck associated with typical analytical procedures such as HPLC analysis. In this thesis methodologies were explored to reduce the error associated with multicomponent isotherm studies as well as explore alternative rapid analytical methods for their study.

It was found that the resin volume used in order to study single component isotherms was critical in order to generate isotherms with reduced associated error. 2 µL of chromatography resin was associated with systematic overestimation in adsorbed concentration when elucidating isotherm of moderate interaction strength. Additionally such small resin volumes were associated with increased random error. It was found that the results generated using 20 and 50 µL of resin were in agreement with one another and were not associated with systematic mis-estimation. Error propagation showed that these increased resin volumes were associated with reduced random error. Additionally, error propagation also showed these moderately interacting isotherms were associated with decreased random error when the adsorbed concentration was taken as q^*, rather than the traditional method q. It was also found that the increased protein stock concentration

associated with studies using larger resin volumes affected liquid handling performance, which was subsequently optimised.

In order to study multicomponent isotherms an analytical method which can ascertain the titre of individual protein concentrations within a mixture is required. Extensive literature review found that a single traditional HPLC method would not resolve the mixture of 3 proteins to baseline and an additional deconvolution method would be required. An alternative rapid HT analytical method was instead used to ascertain the concentration of individual proteins within a mixture. The method involved the collection of UV spectra which were used to build a multivariate statistical model, which was used to return estimates of protein concentration. Such methods have been used before in the quantification of proteins (Brestrich et al., 2014; Dismer et al., 2013; Hansen et al., 2013), and in the study of binary adsorption isotherms (Baumann et al., 2016). To our knowledge this is the first time such a method has been used for this application where none of the proteins contain a chromophore and for proteins of such size, which such small separation factors. In this work the assayable ranges of the analytical method were increased which also improves the sensitivity of the assay. Additionally, the use of local multivariate models in regions where quantification becomes difficult due to the low proportion of signal for one of the components has not been reported before. Moreover, models were calibrated using an over-populated space filling grid layout meaning performance of the method in any region was well understood, unlike in previously reported models where designs using minimal sample numbers were used. It was found that application of the UV spectra-multivariate method was more error prone for ternary quantification in comparison to binary. This is likely because the ternary mixture of proteins chosen is considered to be spectrally quite similar which underlines a limitation to this approach (Hansen et al., 2013).

The HT plate platform optimised for isotherm determination was used in conjunction with the rapid HT analytical method in order to generate 3 binary isotherms across different pH and NaCl levels. Propagation of systematic errors carried forward from the UV spectra-multivariate method were quantified in the binary isotherms as well as the random errors which were introduced from various sources. It was found that adsorbed concentration calculated using the traditional mass balance method where the liquid phase solute

concentration before and after equilibration with the sorbent is subtracted (referred to as *q*) is associated with fairly constant random error across different NaCl levels. Measurement of adsorbed concentration using an alternative method, where the adsorbed concentration is eluted and measured directly (know as *q**) was associated with random error which increased with adsorbed concentration. In conditions of moderate or weak interaction strength there is less random error associated measuring adsorbed concentration using *q** rather than *q*. The comparison of adsorbed concentration using *q* and *q** measures was very good for single component isotherms and good in 25/30 examples of binary isotherms. This is a strong indicator that the methodological approach is robust. Additionally, the use of rapid high throughput analytics not only facilitates the study of multicomponent isotherms in terms of speeding up such studies but it always allows more data to be collected in the system improving data reliability. The quantification of *q* and *q** measures of adsorption would not be realistic using traditional HPLC methods which are comparatively slow. Moreover, the rapid method allowed multiple dilution levels of each sample to be analysed and outlier rejection methodologies to be applied to the different version of data points improving data reliability.

The application of these methodologies to ternary protein isotherms, which historically have received very little attention (Close et al., 2014; Melter et al., 2007; Tao et al., 2011), was more problematic. There was generally poor agreement between *q* and *q** measures of adsorbed concentration. This is due to the comparatively poorer performance of the ternary PLS models in comparison to the binary models. It was found again that the *q** measure of adsorbed concentration was less error prone in more weakly interacting conditions. It was also found that restrictions on the volume of resin deployed in the slurry plate limited exploration of saturation capacity under certain conditions. These restrictions were the product of what was found when optimising the slurry plate to decrease the systematic and random error.

Once the multicomponent isotherms were elucidated they were then fit to different isotherm formalisms such as the Fruendlich-Langmuir-Jovanovich, steric mass action (SMA) and Padé descriptions. Whilst none of the formalisms found were able to properly describe

the complex behaviours seen in data of such high granularity it was found that the SMA and Padé isotherms were the worst performers and should be de-emphasised during such studies.

Because none of the formalisms could adequately describe the complex behaviours seen in the multicomponent isotherm it was not possible to run an in silico optimisation of BSA-Ova mixtures. Failing that a heuristic optimisation was rapidly achieved using the UV-spectra method to quantify fractions. There was the appearance of an artefact using this analytical method which manifested as a BSA pre-peak before the elution of Ova. These BSA pre-peaks were successfully identified as artefacts using AEX-HPLC identity analysis. This combination of methods still allowed for rapid analysis and optimisation as fractions in question were pooled meaning multiple preparative runs could be checked for artefact presence in one session of HPLC analysis.

In conclusion, experimental factors were successfully identified which are critical to obtaining high quality data on slurry plate formats. This included the use of an alternative measure of adsorbed concentration where adsorbed protein is eluted and assayed directly rather than via mass balance calculation. Rapid HT analytics were successfully developed and applied to binary isotherms, although their application to ternary isotherms had limited success.

The studies here make the study of binary isotherms a more feasible prospect as methods developed allow the generation of high quality data at a range of interaction strengths which has been problematic previously. Rapid HT analysis on conventional non-chromophore containing proteins will facilitate the study of binary isotherms on industrially relevant therapeutic proteins as well as the possibility of improving data quality by allowing the quantification of equilibrium and elution fractions at multiple dilution levels. Error propagation on multicomponent isotherms showed why the quantification of adsorbed concentration via elution can be favourable under conditions of moderate or weak interaction as well as suggesting why the elucidation of multicomponent isotherms across different NaCl levels can be challenging using traditional methods.

10.2 Future work

During the generation of ternary isotherms it was found that the saturation capacity could not be fully explored for certain load combinations. Other experimental constraints on minimum resin volume and the limited availability of commercial plates with different resin volumes limited such studies. It was suggested that use of 15 µL of Capto^{TM} Q would allow better exploration of the binding capacity in ternary isotherm elucidation whilst keeping systematic and random error under reasonable control. The development of procedures which would allow plates to be packed in house with any desired volume would be very useful for this purpose.

Ternary isotherms could not be successfully elucidated in the studies executed here due to the presence of systematic error in the UV spectra-PLS method. It was suggested that the reason for this difficulty in quantification is due to the spectral similarity of the 3 proteins under study. The application of these methods to a 3 component mixture with less spectral similarity could show that the these methodologies can be applied to high order mixtures given sufficient spectral difference. However, the choice of 3 non-spectrally similar proteins in a real world problem is unlikely. Additionally, the application of AEX-HPLC analysis on the ternary isotherm in conjunction with an optimised deconvolution method would also demonstrate the applicability of this slurry plate methodology to higher order systems. An interesting alternative would be to develop methods for capillary electrochromatography which might be able to separate these proteins to baseline without the need for a deconvolution method (X. Huang et al., 1999).

The identification of a numerical form which is able to capture the complex trends seen in these multicomponent isotherms would allow these isotherms to be used for *in silico* separation optimisation. No such numerical form could be identified here. The use of p-splines which are splines able to fit complex behaviour across multiple dimensions whilst remaining monotonic in behaviour is a tantalising possibility (Bollaerts, Eilers, & van Mechelen, 2006). However, there is currently no accessible toolbox for such an application which could be applied to the multidimensional problem of multicomponent isotherms, the

development of such a toolbox and the application to this problem could be an extremely powerful solution to the well-recognised problem of multi component isotherm fitting.

References

Abdehagh, N., Tezel, F. H., & Thibault, J. (2016). Multicomponent adsorption modeling: isotherms for ABE model solutions using activated carbon F-400. *Adsorption*, *22*(3), 357–370. http://doi.org/10.1007/s10450-016-9784-y

Aboudzadeh, M. R., Aboudzadeh, N., Jiawen, Z., & Bin, W. (2007). Binary protein adsorption to DEAE sepharose FF. *Korean Journal of Chemical Engineering*, *24*(4), 641–647. http://doi.org/10.1007/s11814-007-0017-7

Baumann, P., Huuk, T., Hahn, T., Osberghaus, A., & Hubbuch, J. (2016). Deconvolution of high-throughput multicomponent isotherms using multivariate data analysis of protein spectra. *Engineering in Life Sciences*, *16*(2), 194–201. http://doi.org/10.1002/elsc.201400243

Bensch, M., Schulze Wierling, P., von Lieres, E., & Hubbuch, J. (2005). High Throughput Screening of Chromatographic Phases for Rapid Process Development. *Chemical Engineering & Technology*, *28*(11), 1274–1284. http://doi.org/10.1002/ceat.200500153

Bergander, T. J., Karf, L., & Brannstrom-Carlsson, K. (2008). BIOT 214-Quality aspects of microtiter plate workflow in the screening of chromatographic conditions. *ABSTRACTS OF PAPERS OF THE AMERICAN CHEMICAL SOCIETY*, *236*.

Bergander, T., Nilsson-Valimaa, K., Oberg, K., & Lacki, K. M. (2008). High-Throughput Process Development: Determination of Dynamic Binding Capacity Using Microtiter Filter Plates Filled with Chromatography Resin. *Biotechnology Progress*, *24*(3), 632–639. http://doi.org/10.1021/bp0704687

Bhambure, R., & Rathore, a. S. (2013). Chromatography process development in the quality by design paradigm I: Establishing a high-throughput process development platform as a tool for estimating “characterization space” for an ion exchange chromatography step. *Biotechnology Progress*, *29*(2), 403–414. http://doi.org/10.1002/btpr.1705

Bollaerts, K., Eilers, P. H. C., & van Mechelen, I. (2006). Simple and multiple P-splines regression with shape constraints. *The British Journal of Mathematical and Statistical Psychology*, *59*(Pt 2), 451–469. http://doi.org/10.1348/000711005X84293

Brestrich, N., Briskot, T., Osberghaus, A., & Hubbuch, J. (2014). A tool for selective inline quantification of co-eluting proteins in chromatography using spectral analysis and partial least squares regression. *Biotechnology and Bioengineering*, *111*(7), 1365–1373. http://doi.org/10.1002/bit.25194

Buchacher, A., & Iberer, G. (2006). Purification of intravenous immunoglobulin G from human plasma – aspects of yield and virus safety. *Biotechnology Journal*, *1*(2), 148–163. http://doi.org/10.1002/biot.200500037

Cano, T., Offringa, N. D., & Willson, R. C. (2005). Competitive ion-exchange adsorption of proteins: Competitive isotherms with controlled competitor concentration. *Journal of Chromatography A*, *1079*(1–2), 116–126. http://doi.org/10.1016/j.chroma.2005.03.120

Carr, D. (2002). The Handbook of Analysis and Purification of Peptides and Proteins by Reversed-Phase HPLC.

Carta, G., & Jungbauer, A. (2010). Adsorption Equilibria. In *Protein Chromatography* (pp. 145–160). Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA. http://doi.org/10.1002/9783527630158.ch5

Chen, W.-D., Hu, H.-H., & Wang, Y.-D. (2006). Analysis of steric mass-action model for protein adsorption equilibrium onto porous anion-exchange adsorbent. *Chemical Engineering Science*, *61*(21), 7068–7076. http://doi.org/10.1016/j.ces.2006.07.036

Chhatre, S., & Titchener-Hooker, N. J. (2009). Review: Microscale methods for high-throughput chromatography development in the pharmaceutical industry. *Journal of Chemical Technology & Biotechnology*, *84*(7), 927–940. http://doi.org/10.1002/jctb.2125

Chilamkurthi, S., Willemsen, J.-H., van der Wielen, L. A. M., Poiesz, E., & Ottens, M. (2012). High-throughput determination of adsorption equilibria for chromatographic oligosaccharide separations. *Journal of Chromatography A*, *1239*, 22–34. http://doi.org/10.1016/j.chroma.2012.03.042

Close, E. J., Salm, J. R., Bracewell, D. G., & Sorensen, E. (2014). Modelling of industrial biopharmaceutical multicomponent chromatography. *Chemical Engineering Research and Design*, *92*(7), 1304–1314. http://doi.org/10.1016/j.cherd.2013.10.022

Coffman, J. L., Kramarczyk, J. F., & Kelley, B. D. (2008). High-throughput screening of chromatographic separations: I. Method development and column modeling. *Biotechnology and Bioengineering*, *100*(4), 605–618. http://doi.org/10.1002/bit.21904

Conder, J. R., & Young, C. L. (1979). *Physicochemical Measurements by Gas Chromatography*. New York: John Wiley & Sons.

Daly, A. E., Gilar, M., Gebler, J. C., Corporation, W., & R, L. S. (2003). Investigation of Protein Recovery and Memory Effects in Reversed-Phase and Ion-Exchange Chromatography.

Demin, A. a., Mogilevskaya, A. D., & Samsonov, G. V. (1997). Synergistic effects in the processes of protein multicomponent sorption. *Journal of Chromatography A*, *760*(1), 105–115. http://doi.org/10.1016/S0021-9673(96)00825-4

Di Marco, V. B., & Bombi, G. G. (2001). Mathematical functions for the representation of chromatographic peaks. *Journal of Chromatography A*, *931*(1–2), 1–30. http://doi.org/10.1016/S0021-9673(01)01136-0

Dimitrov, D. (2012). *Therapeutic Proteins*. (V. Voynov & J. A. Caravella, Eds.) (Vol. 899). Totowa, NJ: Humana Press. http://doi.org/10.1007/978-1-61779-921-1

Dismer, F., Hansen, S., Oelmeier, S. A., & Hubbuch, J. (2013). Accurate retention time determination of co-eluting proteins in analytical chromatography by means of spectral data. *Biotechnology and Bioengineering*, *110*(3), 683–693.

http://doi.org/10.1002/bit.24738

Ecker, D. M., Jones, S. D., & Levine, H. L. (2015). The therapeutic monoclonal antibody market. *mAbs*, *7*(1), 9–14. http://doi.org/10.4161/19420862.2015.989042

Eriksson, L., Byrne, T., Johansson, E., Tyrgg, J., & Vikström. (2013). *Multi- and Megavariate Data Analysis Basic Principles and Applications* (3rd ed.). Mälmo: MKS Umetrics.

Ersson, B., Rydén, L., & Janson, J.-C. (2011). Introduction to Protein Purification. In *Protein Purification: Principles, High Resolution Methods, and Applications: Third Edition* (pp. 1–22). http://doi.org/10.1002/9780470939932.ch1

Faraji, N., Zhang, Y., & Ray, A. K. (2015). Determination of adsorption isotherm parameters for minor whey proteins by gradient elution preparative liquid chromatography. *Journal of Chromatography A*, *1412*, 67–74. http://doi.org/10.1016/j.chroma.2015.08.004

Fargues, C., Bailly, M., & Grevillot, G. (1998). Adsorption of BSA and Hemoglobin on Hydroxyapatite Support: Equilibria and Multicomponent Dynamic Adsorption. *Adsorption*, *4*(1), 5–16. http://doi.org/10.1023/A:1008822918494

Fausnaugh, J. L., Kennedy, L. A., & Regnier, F. E. (1984). Comparison of hydrophobic-interaction and reversed-phase chromatography of proteins. *Journal of Chromatography A*, *317*, 141–155. http://doi.org/10.1016/S0021-9673(01)91654-1

Felton, M. (2004). Liquid Handling: Dispensing Reliability. *Today's Chemist at Work*, 51–54. http://doi.org/10.1038/nbt1105-1445b

Finette, G. M. S., Baharin, B. S., Mao, Q.-M., & Hearn, M. T. W. (1997). Adsorption Behavior of Multicomponent Protein Mixtures Containing α1-Proteinase Inhibitor with the Anion Exchanger, 2-(Diethylamino)ethyl-Spherodex. *Biotechnology Progress*, *13*(3), 265–275. http://doi.org/10.1021/bp9700059

Fregeau, C., Yensen, C., Elliott, J., & Fourney, R. (2007). Optimized Configuration of Fixed-Tip

Robotic Liquid-Handling Stations for the Elimination of Biological Sample Cross-Contamination. *Journal of the Association for Laboratory Automation*, *12*(6), 339–354. http://doi.org/10.1016/j.jala.2007.08.001

Fritz, W., & Schluender, E.-U. (1974). Simultaneous adsorption equilibria of organic solutes in dilute aqueous solutions on activated carbon. *Chemical Engineering Science*, *29*(5), 1279–1282. http://doi.org/10.1016/0009-2509(74)80128-4

Garje, G., Hartmann, R., Papamicmichael, N., Deckwer, W.-D., & Anspach, F. B. (1999). The Influence of Protein Size on Adsorption Kinetics and Equilibria in Ion-Exchange Chromatography. *Separation Science and Technology*, *34*(13), 2521–2538. http://doi.org/10.1081/SS-100100788

Golshan-Shirazi, S., & Guiochon, G. (1994). Modeling of preparative liquid chromatography. *Journal of Chromatography A*, *658*(2), 149–171. http://doi.org/10.1016/0021-9673(94)80013-8

Gu, H., & Deng, Y. (2007). Dilution Effect in Multichannel Liquid-Handling System Equipped with Fixed Tips: Problems and Solutions for Bioanalytical Sample Preparation. *Journal of the Association for Laboratory Automation*, *12*(6), 355–362. http://doi.org/10.1016/j.jala.2007.07.002

Guiochon, G., Felinger, A., & Shirazi, D. G. (2006). *Fundamentals of Preparative and Nonlinear Chromatography*. Academic Press.

Guiochon, G., & Katti, A. (1987). Preparative liquid chromatography. *Chromatographia*, *24*(1), 165–189. http://doi.org/10.1007/BF02688479

Gutsche, C. D. (1970). Catalysis in Chemistry and Enzymology. William P. Jencks. McGraw-Hill, New York, 1969. xvi + 656 pp., illus. $14.50. McGraw-Hill Series in Advanced Chemistry. *Science*, *168*(3935), 1080–1081. http://doi.org/10.1126/science.168.3935.1080-a

Hagel, L. (2011). Gel Filtration: Size Exclusion Chromatography. In *Protein Purification :*

Principles, High Resolution Methods, and Applications (Vol. V, pp. 51–91). http://doi.org/10.1002/9780470939932.ch3

Hansen, S. K., Jamali, B., & Hubbuch, J. (2013). Selective high throughput protein quantification based on UV absorption spectra. *Biotechnology and Bioengineering*, *110*(2), 448–460. http://doi.org/10.1002/bit.24712

Hansen, S. K., Skibsted, E., Staby, A., & Hubbuch, J. (2011). A label-free methodology for selective protein quantification by means of absorption measurements. *Biotechnology and Bioengineering*, *108*(11), 2661–2669. http://doi.org/10.1002/bit.23229

Herrmann, T., Schroder, M., & Hubbuch, J. (2006). Generation of Equally Sized Particle Plaques Using Solid-Liquid Suspensions. *Biotechnology Progress*, *22*(3), 914–918. http://doi.org/10.1021/bp050296i

Hjertén, S., Yao, K., Eriksson, K., & Johansson, B. (1986). Gradient and isocratic high-performance hydrophobic interaction chromatography of proteins on agarose columns. *Journal of Chromatography A*, *359*, 99–109. http://doi.org/10.1016/0021-9673(86)80065-6

Huang, J.-X., & Horváth, C. (1987). Adsorption isotherms on high-performance liquid chromatographic sorbents. *Journal of Chromatography A*, *406*, 275–284. http://doi.org/10.1016/S0021-9673(00)94035-4

Huang, X., Zhang, J., & Horváth, C. (1999). Capillary electrochromatography of proteins and peptides with porous-layer open-tubular columns. *Journal of Chromatography A*, *858*(1), 91–101. http://doi.org/10.1016/S0021-9673(99)00795-5

Jacobson, J. M., Frenz, J. H., & Horvath, C. G. (1987). Measurement of competitive adsorption isotherms by frontal chromatography. *Industrial & Engineering Chemistry Research*, *26*(1), 43–50. http://doi.org/10.1021/ie00061a009

Janson, J.-C., & Jönsson, J. Å. (2011). Introduction to Chromatography. In *Protein Purification :*

Principles, High Resolution Methods, and Applications (pp. 23–50). http://doi.org/10.1002/9780470939932.ch2

Kalambet, Y., Kozmin, Y., Mikhailova, K., Nagaev, I., & Tikhonov, P. (2011). Reconstruction of chromatographic peaks using the exponentially modified Gaussian function. *Journal of Chemometrics*, *25*(7), 352–356. http://doi.org/10.1002/cem.1343

Kamga, M.-H., Woo Lee, H., Liu, J., & Yoon, S. (2013). Quantification of protein mixture in chromatographic separation using multi-wavelength UV spectra. *Biotechnology Progress*, *29*(3), 664–671. http://doi.org/10.1002/btpr.1712

Karlsson, E., & Hirsh, I. (2011). Ion Exchange Chromatography. In *Protein Purification: Principles, High Resolution Methods, and Applications* (pp. 93–133). http://doi.org/10.1002/9780470939932.ch4

Katiyar, A., Ji, L., Smirniotis, P., & Pinto, N. G. (2005). Protein adsorption on the mesoporous molecular sieve silicate SBA-15: effects of pH and pore size. *Journal of Chromatography A*, *1069*(1), 119–126. http://doi.org/10.1016/j.chroma.2004.10.077

Kittelmann, J., Ottens, M., & Hubbuch, J. (2015). Robust high-throughput batch screening method in 384-well format with optical in-line resin quantification. *Journal of Chromatography B*, *988*, 98–105. http://doi.org/10.1016/j.jchromb.2015.02.028

Konstantinidis, S., Kong, S., & Titchener-Hooker, N. (2013). Identifying analytics for high throughput bioprocess development studies. *Biotechnology and Bioengineering*, *110*(7), 1924–1935. http://doi.org/10.1002/bit.24850

Kopaciewicz, W., Rounds, M., Fausnaugh, J., & Regnier, F. E. (1983). Retention model for high-performance ion-exchange chromatography. *Journal of Chromatography A*, *266*(9395), 3–21. http://doi.org/10.1016/S0021-9673(01)90875-1

Kumar, P., Lau, P. W., Kale, S., Johnson, S., Pareek, V., Utikar, R., & Lali, A. (2014). Kafirin adsorption on ion-exchange resins: Isotherm and kinetic studies. *Journal of*

Chromatography A, *1356*, 105–116. http://doi.org/10.1016/j.chroma.2014.06.035

Lacki, K. M., & Brekkan, E. (2011). High Throughput Screening Techniques in Protein Purification. In *Protein Purification: Principles, High …* (pp. 487–506). http://doi.org/10.1002/9780470939932.ch20

Lagarias, J. C., Reeds, J. A., Wright, M. H., & Wright, P. E. (1998). Convergence Properties of the Nelder--Mead Simplex Method in Low Dimensions. *SIAM Journal on Optimization*, *9*(1), 112–147. http://doi.org/10.1137/S1052623496303470

LeVan, M. D., & Vermeulen, T. (1981). Binary Langmuir and Freundlich isotherms for ideal adsorbed solutions. *The Journal of Physical Chemistry*, *85*(22), 3247–3250. http://doi.org/10.1021/j150622a009

Lewus, R. K., & Carta, G. (1999). Binary protein adsorption on gel-composite ion-exchange media. *AIChE Journal*, *45*(3), 512–522. http://doi.org/10.1002/aic.690450308

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. http://doi.org/10.1016/j.jesp.2013.03.013

Liang, J., Fieg, G., & Jakobtorweihen, S. (2015). Ion-Exchange Adsorption of Proteins: Experiments and Molecular Dynamics Simulations. *Chemie Ingenieur Technik*, *87*(7), 903–909. http://doi.org/10.1002/cite.201400095

Liang, J., Fieg, G., Shi, Q.-H., & Sun, Y. (2012). Single and binary adsorption of proteins on ion-exchange adsorbent: The effectiveness of isothermal models. *Journal of Separation Science*, *35*(17), 2162–2173. http://doi.org/10.1002/jssc.201200101

Lisec, O., Hugo, P., & Seidel-Morgenstern, A. (2001). Frontal analysis method to determine competitive adsorption isotherms. *Journal of Chromatography A*, *908*(1–2), 19–34. http://doi.org/10.1016/S0021-9673(00)00966-3

Melter, L., Ströhlein, G., Butté, A., & Morbidelli, M. (2007). Adsorption of monoclonal antibody variants on analytical cation-exchange resin. *Journal of Chromatography A*, *1154*(1–2), 121–131. http://doi.org/10.1016/j.chroma.2007.03.130

Nfor, B. K., Noverraz, M., Chilamkurthi, S., Verhaert, P. D. E. M., van der Wielen, L. A. M., & Ottens, M. (2010). High-throughput isotherm determination and thermodynamic modeling of protein adsorption on mixed mode adsorbents. *Journal of Chromatography A*, *1217*(44), 6829–6850. http://doi.org/10.1016/j.chroma.2010.07.069

Osberghaus, A. Z. (2012). *Optimization of chromatographic multi-component separations in silico using HTS-data*. Karlsruhe Instituts fur Technologie.

Osberghaus, a., Baumann, P., Hepbildikler, S., Nath, S., Haindl, M., von Lieres, E., & Hubbuch, J. (2012). Detection, Quantification, and Propagation of Uncertainty in High-Throughput Experimentation by Monte Carlo Methods. *Chemical Engineering & Technology*, *35*(8), 1456–1464. http://doi.org/10.1002/ceat.201100610

Pettersson, S. W. (2011). High-Resolution Reversed-Phase Chromatography of Proteins. In *Protein Purification: Principles, High Resolution Methods, and Applications: Third Edition* (pp. 135–164). http://doi.org/10.1002/9780470939932.ch5

Price, K., Storn, R., & Lampinen, J. (2005). *Differential Evolution*. Berlin/Heidelberg: Springer-Verlag. http://doi.org/10.1007/3-540-31306-0

Rathore, A. S., Li, X., Bartkowski, W., Sharma, A., & Lu, Y. (2009). Case study and application of process analytical technology (PAT) towards bioprocessing: Use of tryptophan fluorescence as at-line tool for making pooling decisions for process chromatography. *Biotechnology Progress*, *25*(5), 1433–1439. http://doi.org/10.1002/btpr.212

Reck, J. M., Pabst, T. M., Hunter, A. K., Wang, X., & Carta, G. (2015). Adsorption Equilibrium and Kinetics of Monomer-Dimer Monoclonal Antibody Mixtures on a Cation Exchange Resin. *Journal of Chromatography A*. http://doi.org/10.1016/j.chroma.2015.05.007

Rege, K., Tugcu, N., & Cramer, S. M. (2003). Predicting Column Performance in Displacement Chromatography from High Throughput Screening Batch Experiments. *Separation Science and Technology*, *38*(7), 1499–1517. http://doi.org/10.1081/SS-120019089

Rhode, H., Schulze, M., Renard, S., Zimmermann, P., Moore, T., Cumme, G. a, & Horn, A. (2004). An Improved Method for Checking HTS/uHTS Liquid-Handling Systems. *Journal of Biomolecular Screening*, *9*(8), 726–733. http://doi.org/10.1177/1087057104269496

Roth, C. M., Unger, K. K., & Lenhoff, A. M. (1996). Mechanistic model of retention in protein ion-exchange chromatography. *Journal of Chromatography A*, *726*(1–2), 45–56. http://doi.org/10.1016/0021-9673(95)01043-2

Rousseeuw, P. J., & Hubert, M. (2011). Robust statistics for outlier detection. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*(1), 73–79. http://doi.org/10.1002/widm.2

Schmidt-Traub, H. (2005). *Preparative Chromatography of Fine Chemicals and Pharmaceutical Agents* (1st ed). Weinheim: Wiley-VCH Ver ag GmBH & Co KGaA.

Seidel-Morgenstern, A. (2004). Experimental determination of single solute and competitive adsorption isotherms. *Journal of Chromatography A*, *1037*(1–2), 255–272. http://doi.org/10.1016/j.chroma.2003.11.108

Shi, Q., Zhou, Y., & Sun, Y. (2008). Influence of pH and Ionic Strength on the Steric Mass-Action Model Parameters around the Isoelectric Point of Protein. *Biotechnology Progress*, *21*(2), 516–523. http://doi.org/10.1021/bp049735o

Sjöström, M., Wold, S., Lindberg, W., Persson, J.-Å., & Martens, H. (1983). A multivariate calibration problem in analytical chemistry solved by partial least-squares models in latent variables. *Analytica Chimica Acta*, *150*(C), 61–70. http://doi.org/10.1016/S0003-2670(00)85460-4

Skidmore, G. L., & Chase, H. a. (1990). Two-component protein adsorption to the cation

exchanger S Sepharose FF. *Journal of Chromatography*, *505*(2), 329–47. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2355064

Skidmore, G. L., Hortsmann, B. J., & Chase, H. A. (1990). Modelling single-component protein adsorption to the cation exchanger s sepharose® FF. *Journal of Chromatography A*, *498*(2), 113–128. http://doi.org/10.1016/S0021-9673(01)84240-0

Staby, A., & Jensen, I. H. (2001). Comparison of chromatographic ion-exchange resins. *Journal of Chromatography A*, *908*(1–2), 149–161. http://doi.org/10.1016/S0021-9673(00)00999-7

Staby, A., Jensen, R. H., Bensch, M., Hubbuch, J., Dünweber, D. L., Krarup, J., … Jensen, I. H. (2007). Comparison of chromatographic ion-exchange resins. *Journal of Chromatography A*, *1164*(1–2), 82–94. http://doi.org/10.1016/j.chroma.2007.06.048

Streat, M. (1995). "The Waters Were Made Sweet". Advances in Ion Exchange Technology. *Industrial & Engineering Chemistry Research*, *34*(8), 2841–2848. http://doi.org/10.1021/ie00047a038

Tao, Y., & Carta, G. (2008). Rapid monoclonal antibody adsorption on dextran-grafted agarose media for ion-exchange chromatography. *Journal of Chromatography A*, *1211*(1–2), 70–79. http://doi.org/10.1016/j.chroma.2008.09.096

Tao, Y., Carta, G., Ferreira, G., & Robbins, D. (2011). Adsorption of deamidated antibody variants on macroporous and dextran-grafted cation exchangers: I. Adsorption equilibrium. *Journal of Chromatography A*, *1218*(11), 1519–1529. http://doi.org/10.1016/j.chroma.2011.01.049

The American Society of Mechanical Engineers. (2005). Test uncertainty.

Toueille, M., Uzel, A., Depoisier, J.-F., & Gantier, R. (2011). Designing new monoclonal antibody purification processes using mixed-mode chromatography sorbents. *Journal of Chromatography B*, *879*(13–14), 836–843.

http://doi.org/10.1016/j.jchromb.2011.02.047

Traylor, S. J., Xu, X., Li, Y., Jin, M., & Li, Z. J. (2014). Adaptation of the pore diffusion model to describe multi-addition batch uptake high-throughput screening experiments. *Journal of Chromatography A*, *1368*, 100–106. http://doi.org/10.1016/j.chroma.2014.09.058

Treier, K., Hansen, S., Richter, C., Diederich, P., Hubbuch, J., & Lester, P. (2012). High-throughput methods for miniaturization and automation of monoclonal antibody purification processes. *Biotechnology Progress*, *28*(3), 723–732. http://doi.org/10.1002/btpr.1533

Velayudhan, A., & Horváth, C. (1988). Preparative chromatography of proteins. *Journal of Chromatography A*, *443*, 13–29. http://doi.org/10.1016/S0021-9673(00)94779-4

Viklund, C., Svec, F., Frechet, J. M. J., & Irgum, K. (1997). Fast Ion-Exchange HPLC of Proteins Using Porous Poly(glycidyl methacrylate-co-ethylene dimethacrylate) Monoliths Grafted with Poly(2-acrylamido-2-methyl-1-propanesulfonic acid). *Biotechnology Progress*, *13*(5), 597–600. http://doi.org/10.1021/bp9700667

Waters Corporation. (2017). ACQUITY UPLC Protein BEH SEC Columns and Standards.

Welsh, J. P., Petroff, M. G., Rowicki, P., Bao, H., Linden, T., Roush, D. J., & Pollard, J. M. (2014). A practical strategy for using miniature chromatography columns in a standardized high-throughput workflow for purification development of monoclonal antibodies. *Biotechnology Progress*, *30*(3), 626–635. http://doi.org/10.1002/btpr.1905

Wold, S. (1978). Cross-Validatory Estimation of the Number of Components in Factor and Principal Components Models. *Technometrics*, *20*(4), 397. http://doi.org/10.2307/1267639

Wright, P. R., Muzzio, F. J., & Glasser, B. J. (1998). Batch Uptake of Lysozyme: Effect of Solution Viscosity and Mass Transfer on Adsorption. *Biotechnology Progress*, *14*(6), 913–921. http://doi.org/10.1021/bp980086o

Wu, X.-H., & Lin, B.-C. (2009). Model Modification of Binary Competitive Isotherm. *Journal of Liquid Chromatography & Related Technologies*, *32*(17), 2465–2483. http://doi.org/10.1080/10826070903248254

Xie, I. H., Wang, M. H., Carpenter, R., & Wu, H. Y. (2004). Automated Calibration of TECAN Genesis Liquid Handling Workstation Utilizing an Online Balance and Density Meter. *ASSAY and Drug Development Technologies*, *2*(1), 71–80. http://doi.org/10.1089/154065804322966333

Xu, X., & Lenhoff, A. M. (2009). Binary adsorption of globular proteins on ion-exchange media. *Journal of Chromatography A*, *1216*(34), 6177–6195. http://doi.org/10.1016/j.chroma.2009.06.082

Yamamoto, S., & Ishihara, T. (1999). Ion-exchange chromatography of proteins near the isoelectric points. *Journal of Chromatography A*, *852*(1), 31–36. http://doi.org/10.1016/S0021-9673(99)00593-2

Yang, K. (2016). Dynamic binary protein adsorption in ion-exchange media depicted with a parallel diffusion model derived from Maxwell–Stefan theory. *Chemical Engineering Science*, *139*, 163–172. http://doi.org/10.1016/j.ces.2015.09.027

Zhang, Q., Schimpf, F., Lu, H.-L., Lin, D.-Q., & Yao, S.-J. (2016). Binary Adsorption Processes of Albumin and Immunoglobulin on Hydrophobic Charge-Induction Resins. *Journal of Chemical & Engineering Data*, *61*(3), 1353–1360. http://doi.org/10.1021/acs.jced.5b01108

Zhang, S., & Sun, Y. (2003). A Predictive Model for Salt Effects on the Dye-Ligand Affinity Adsorption Equilibrium of Protein. *Industrial & Engineering Chemistry Research*, *42*(6), 1235–1242. http://doi.org/10.1021/ie020542k

Zhou, X.-P., Su, X.-L., & Sun, Y. (2007). Analysis of the Statistical Thermodynamic Model for Nonlinear Binary Protein Adsorption Equilibria. *Biotechnology Progress*, *23*(5), 0–0. http://doi.org/10.1021/bp070092x

Chapter 11 Additional material

11.1 Appendix A: Additional binary isotherm datasets

Binary isotherms of Ova-Con and BSA-Con were also studied at pH 8 and pH 9 50 mM Tris and at a range of NaCl concentrations. Data sets are presented, systematic and random error propagation in the adsorbed concentration using q and q^* is presented when the isotherm data suggests it is pertinent. BSA-Ova binary isotherm datasets along with the propagation of their error at pH 8 is also presented here. The specific list of binary isotherms presented in this appendix is as follows:

- BSA-Con: pH 9; 0, 25 and 50 mM NaCl, pH 8 0 mM NaCl level
- Ova-Con: pH 9; 0, 25 and 50 mM NaCl, pH 8 0 mM NaCl level
- BSA-Ova: pH 9; 100 mM NaCl pH 8; 0, 50 and 100 mM NaCl

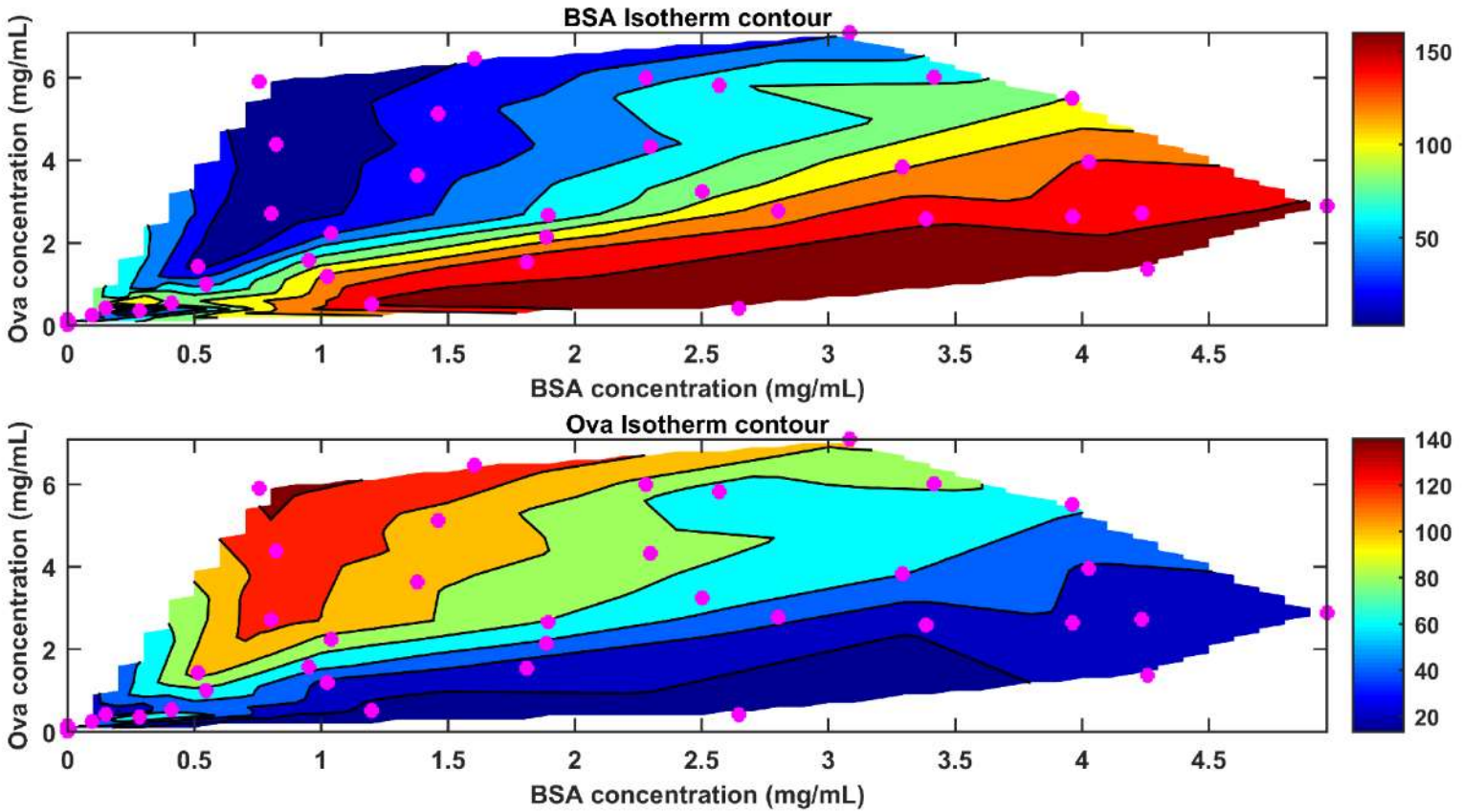

Figure 11-1 Binary isotherm for BSA-Ova pH 9 50 mM NaCl. Isotherm has been drawn as a contour plot as a reference for error propagation plots. x and y axis display average equilibrium concentrations of BSA and Ova measured directly after outlier rejection, contours display adsorbed concentration of BSA in the top panel and Ova in the bottom panel, these have been taken as averages of q and q.*

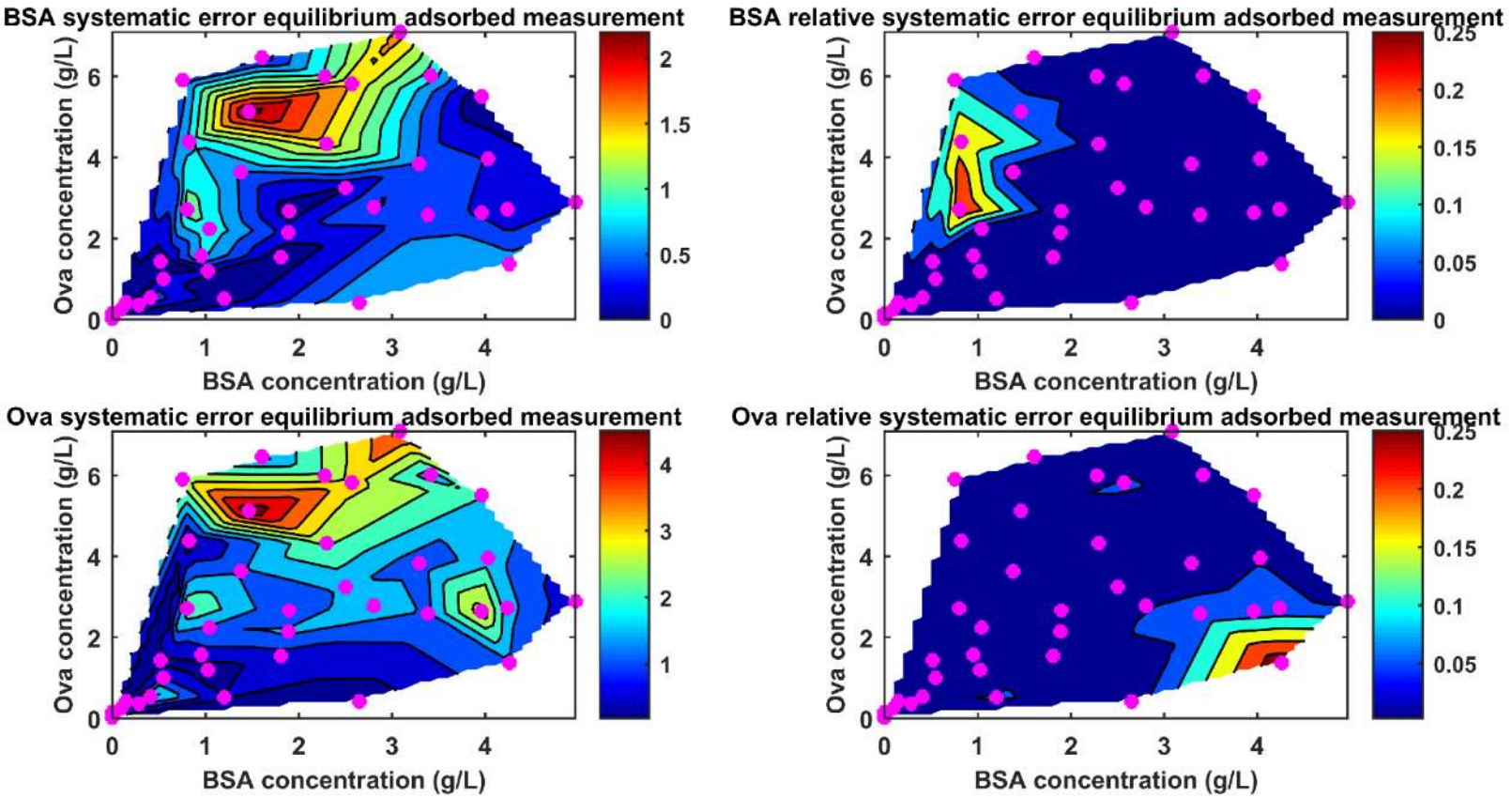

Figure 11-2 Contour plots showing propagated systematic absolute error in adsorbed concentration calculated using q for BSA-Ova binary isotherms at pH 9 50 mM NaCl. Top 2 panels show propagated error for BSA and bottom 2 for Ova. Left hand panels show propagated error in adsorbed concentration mg/mL resin and right hand panels show relative error in adsorbed concentration.

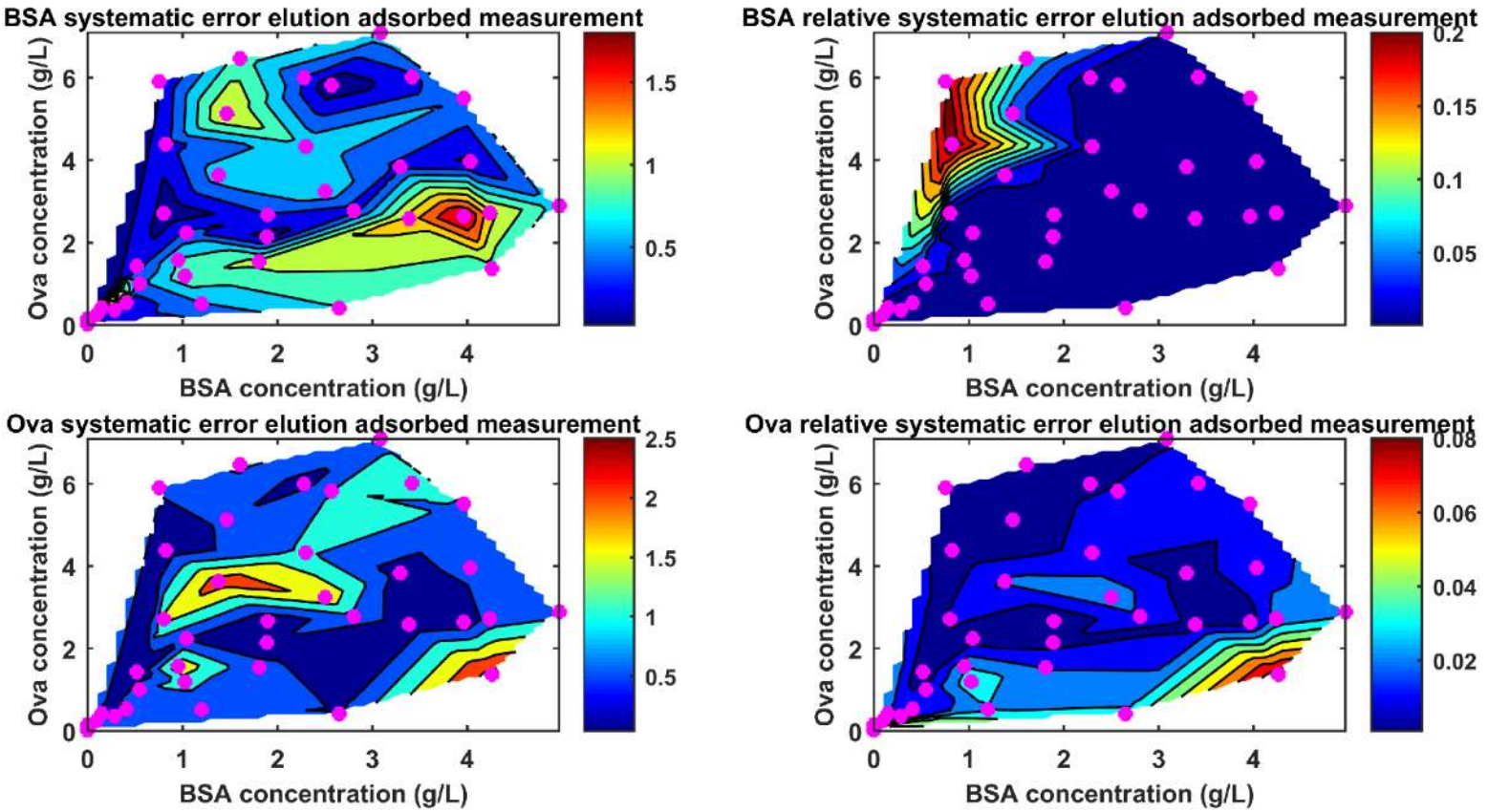

*Figure 11-3 Contour plots showing propagated systematic absolute error in adsorbed concentration calculated using q*for BSA-Ova binary isotherms at pH 9 50 mM NaCl. Top 2 panels show propagated error for BSA and bottom 2 for Ova. Left hand panels show propagated error in adsorbed concentration mg/mL resin and right hand panels show relative error in adsorbed concentration.*

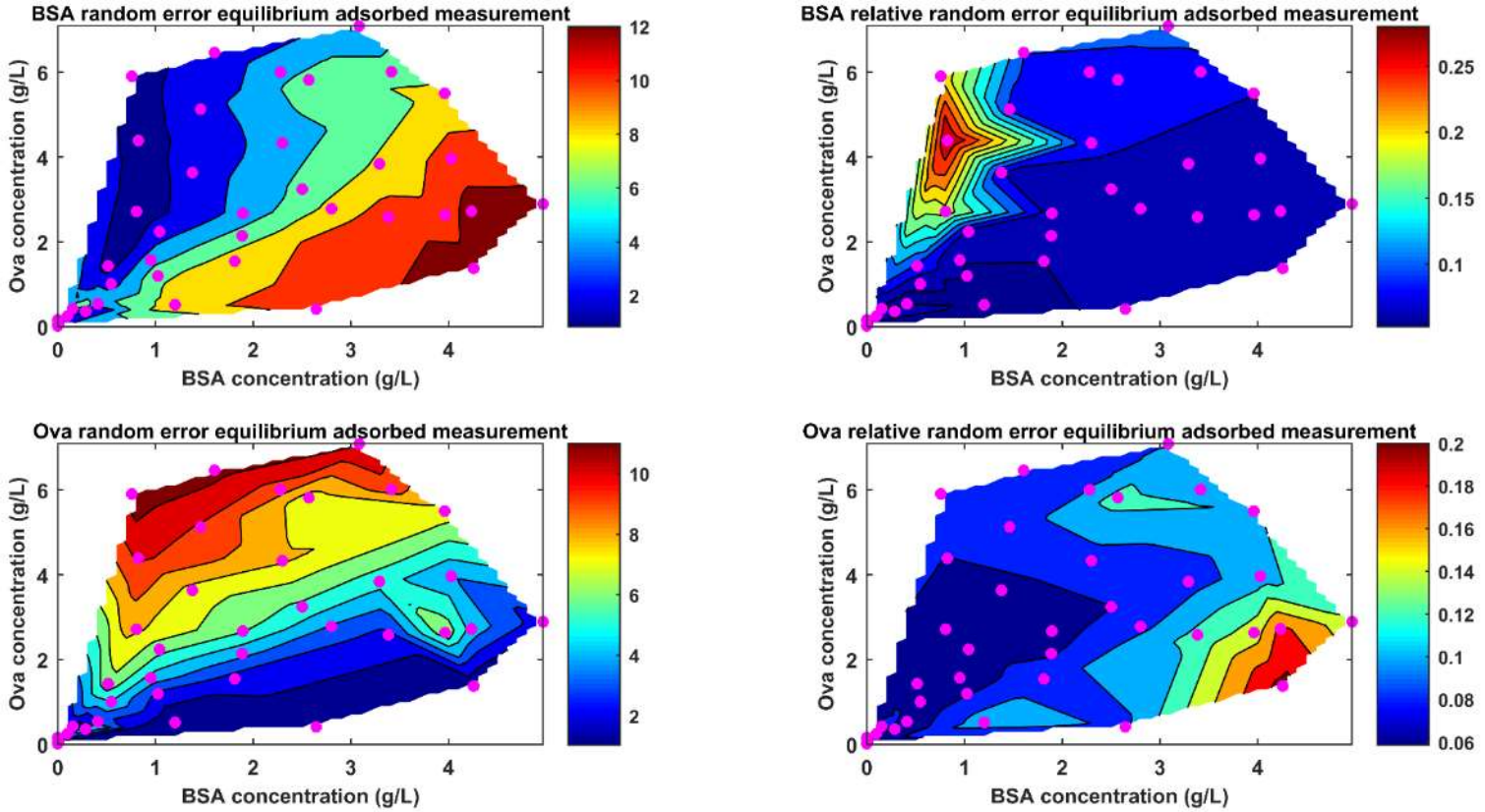

Figure 11-4 Contour plots showing propagated random error in adsorbed concentration calculated using q for BSA-Ova binary isotherms at pH 9 50 mM NaCl. Top 2 panels show propagated error for BSA and bottom 2 for Ova. Left hand panels show propagated error in adsorbed concentration mg/mL resin and right hand panels show relative error in adsorbed concentration.

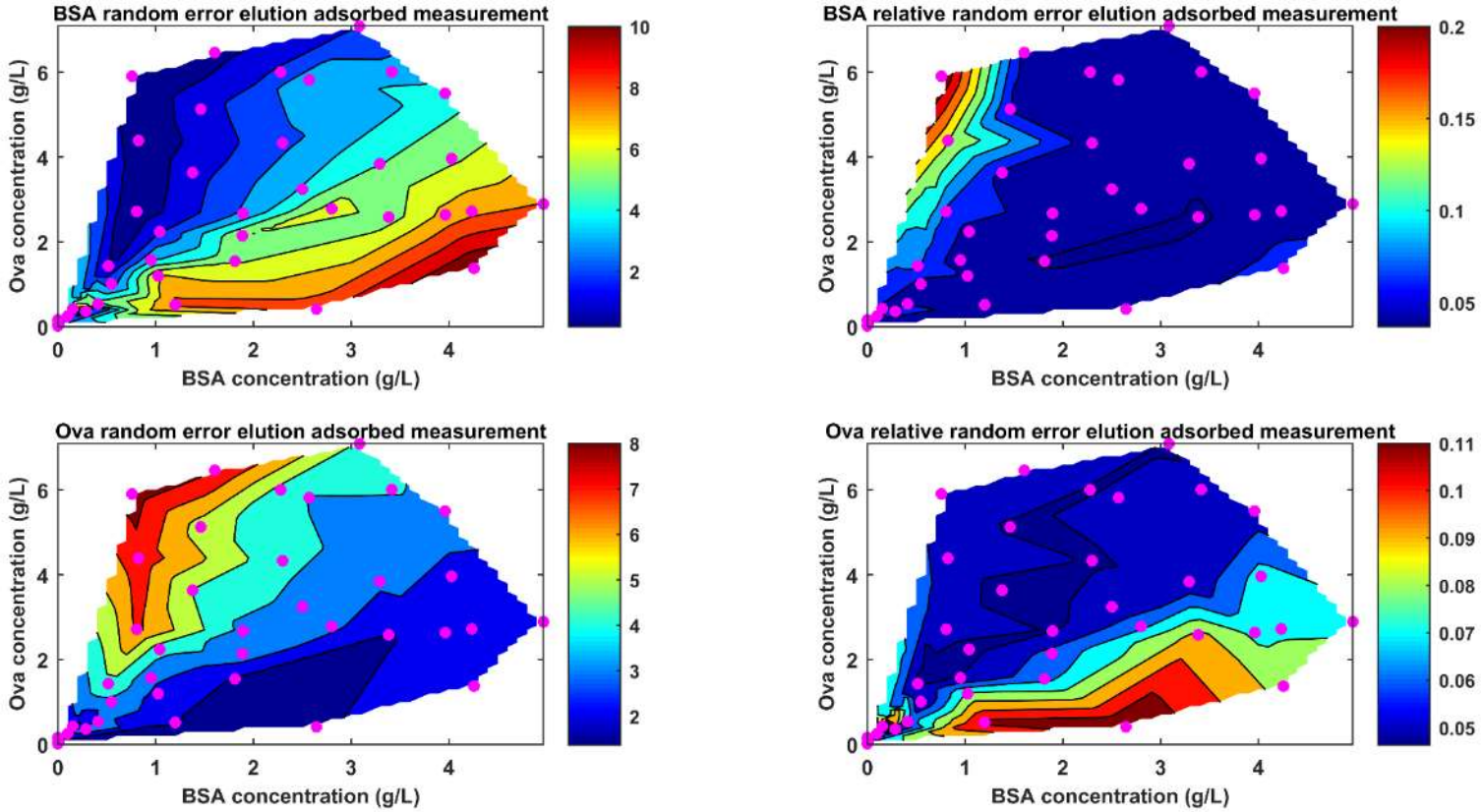

Figure 11-5 Contour plots showing propagated random error in adsorbed concentration calculated using q for BSA-Ova binary isotherms at pH 9 50 mM NaCl. Top 2 panels show propagated error for BSA and bottom 2 for Ova. Left hand panels show propagated error in adsorbed concentration mg/mL resin and right hand panels show relative error in adsorbed concentration.*

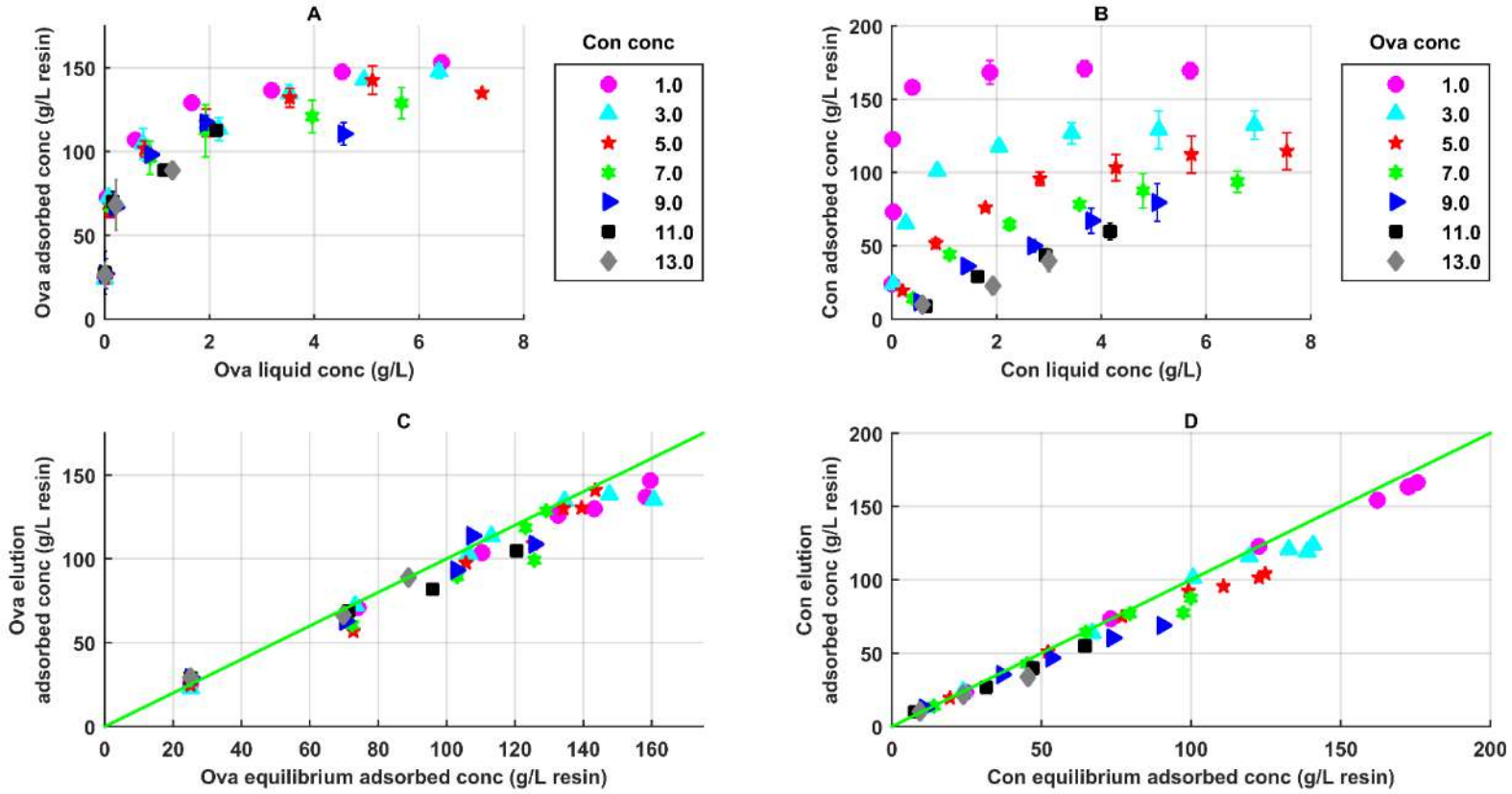

Figure 11-6 Ova-Con binary isotherm pH 9 0 mM NaCl. Panels A & C present Ova liquid and adsorbed concentrations, B & D represent Con data. Panels A and B present the binary isotherms at different starting concentrations of the competitor displayed in the figure legend in mg/mL, the liquid concentration is the average of q and q after outlier rejection. Error bars in panels A and B represent experimental standard deviation. Panels C and D present agreement between q and q*with a line of parity in green.*

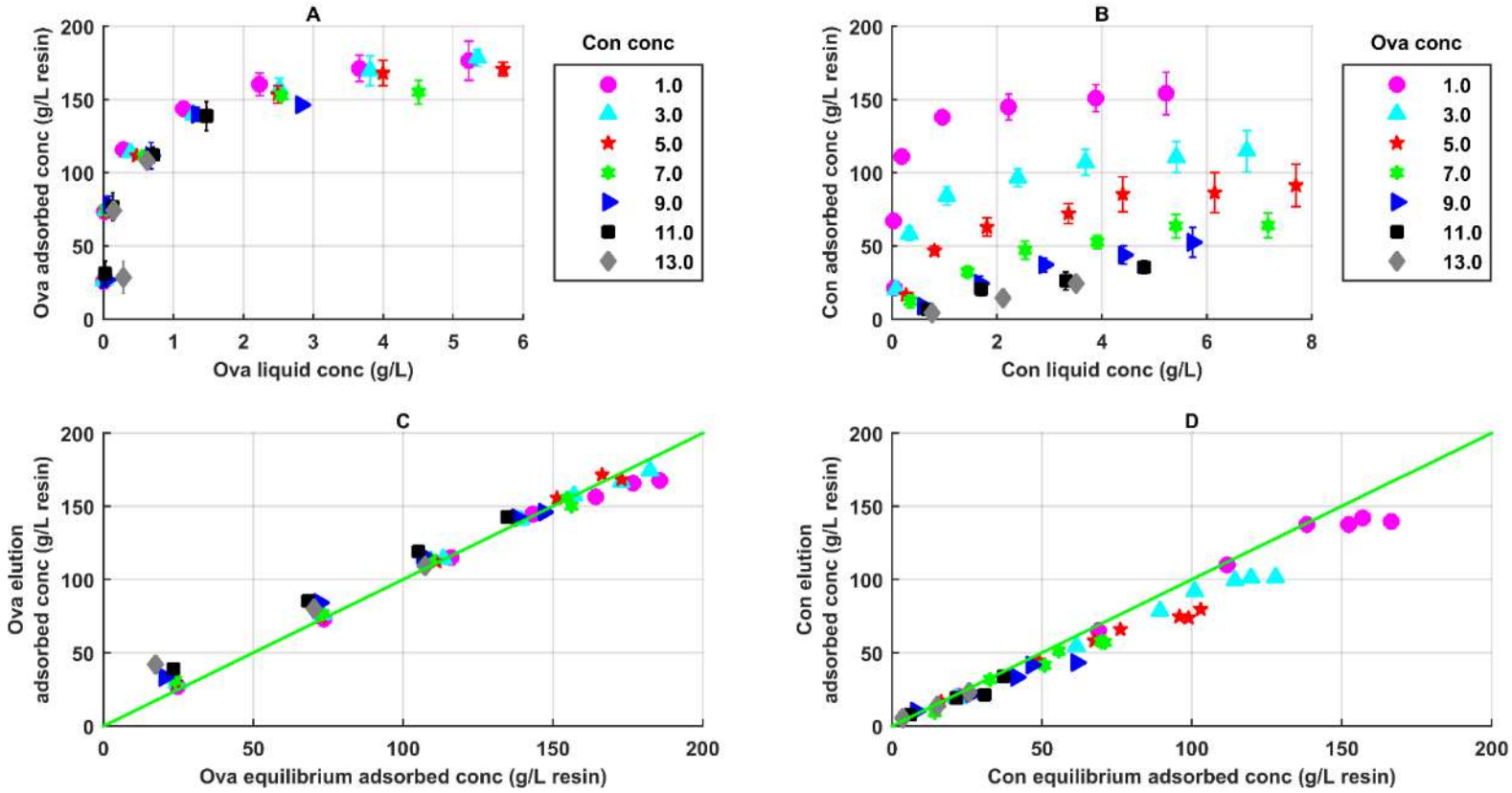

Figure 11-7 Ova-Con binary isotherm pH 9 25 mM NaCl. Panels A & C present Ova liquid and adsorbed concentrations, B & D represent Con data. Panels A and B present the binary isotherms at different starting concentrations of the competitor displayed in the figure legend in mg/mL, the liquid concentration is the average of q and q after outlier rejection. Error bars in panels A and B represent experimental standard deviation. Panels C and D present agreement between q and q*with a line of parity in green. Poor agreement between q and q* is observed for Con at high adsorbed concentration and systematic error is investigated in Figure 11-9 and Figure 11-10.*

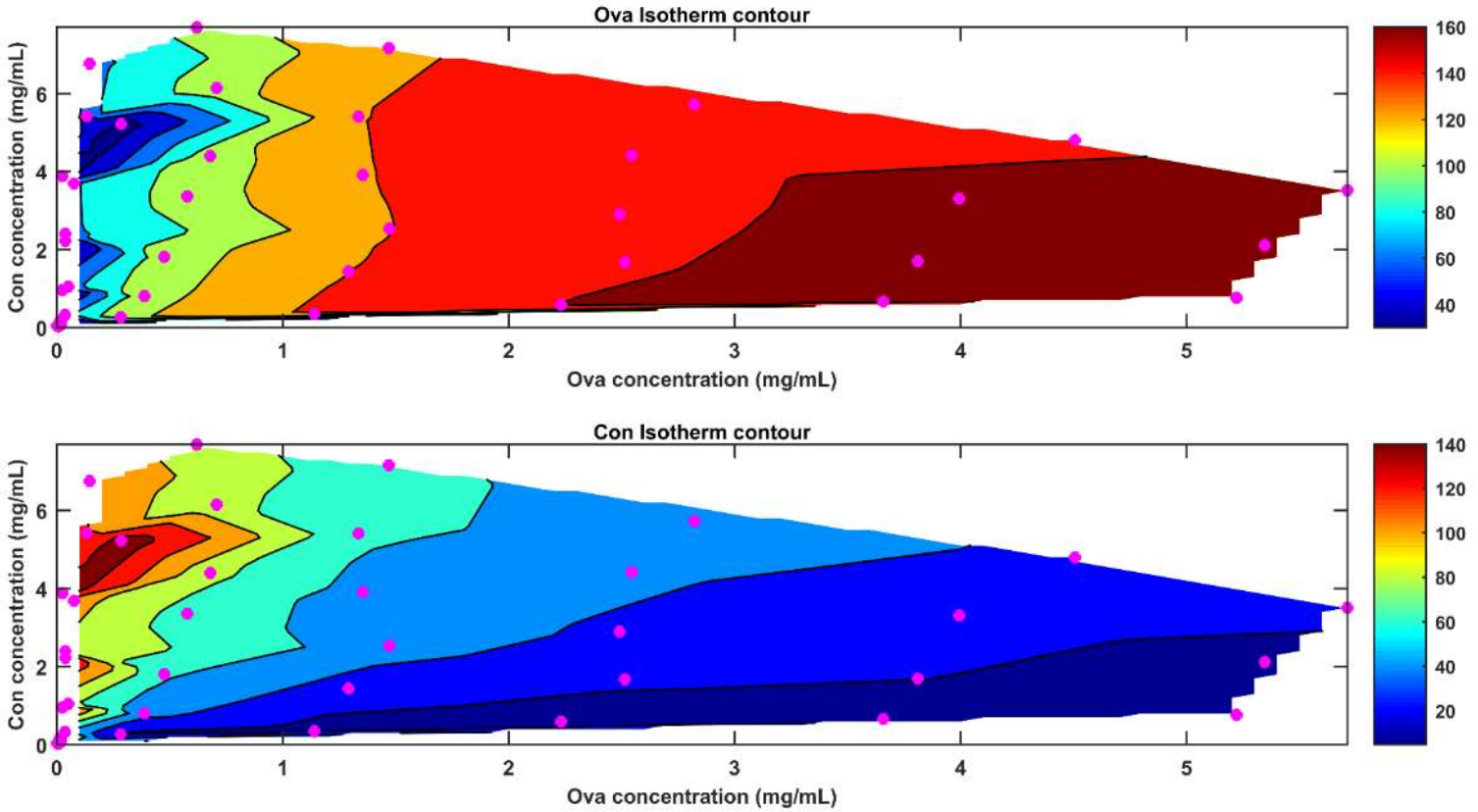

Figure 11-8 Binary isotherm for Ova-Con pH 9 25 mM NaCl. Isotherm has been drawn as a contour plot as a reference for error propagation plots. x and y axis display average equilibrium concentrations of Ova and Con measured directly after outlier rejection, contours display adsorbed concentration of Ova in the top panel and Con in the bottom panel, these have been taken as averages of q and q.*

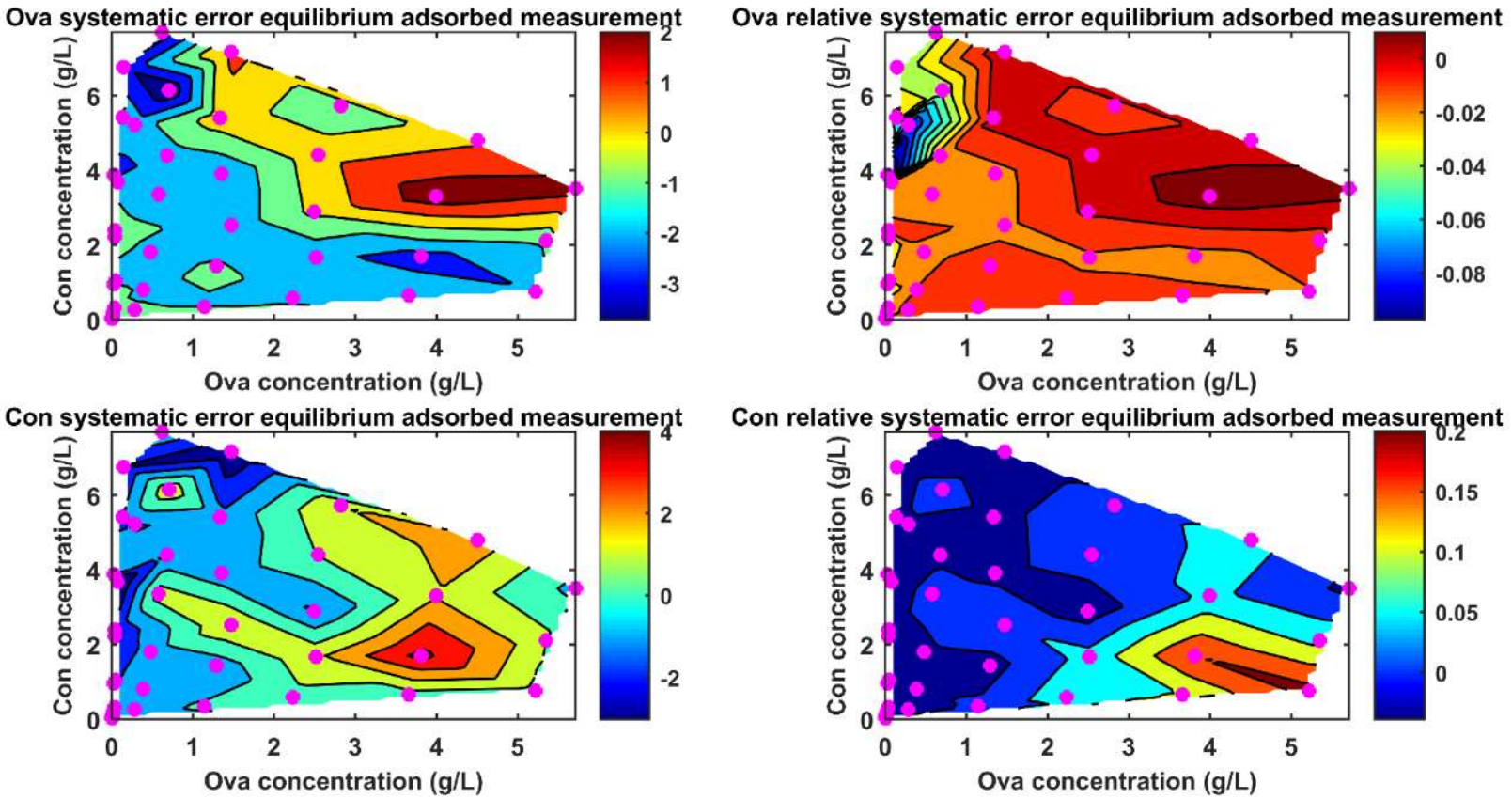

Figure 11-9 Contour plots showing propagated systematic absolute error in adsorbed concentration calculated using q for Ova-Con binary isotherms at pH 9 25 mM NaCl. Top 2 panels show propagated error for Ova and bottom 2 for Con. Left hand panels show propagated error in adsorbed concentration mg/mL resin and right hand panels show relative error in adsorbed concentration.

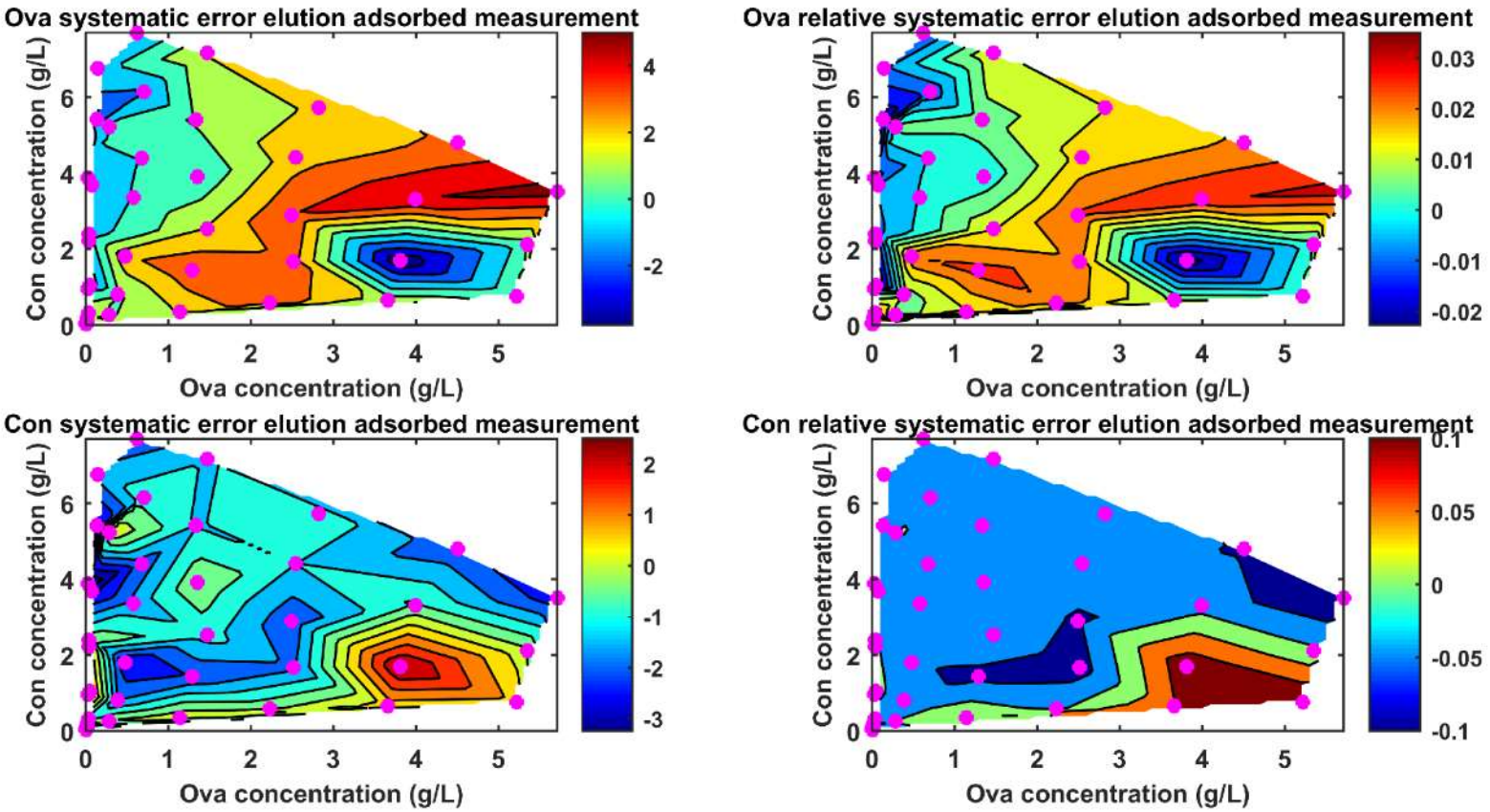

*Figure 11-10 Contour plots showing propagated systematic absolute error in adsorbed concentration calculated using q*for Ova-Con binary isotherms at pH 9 25 mM NaCl. Top 2 panels show propagated error for Ova and bottom 2 for Con. Left hand panels show propagated error in adsorbed concentration mg/mL resin and right hand panels show relative error in adsorbed concentration.*

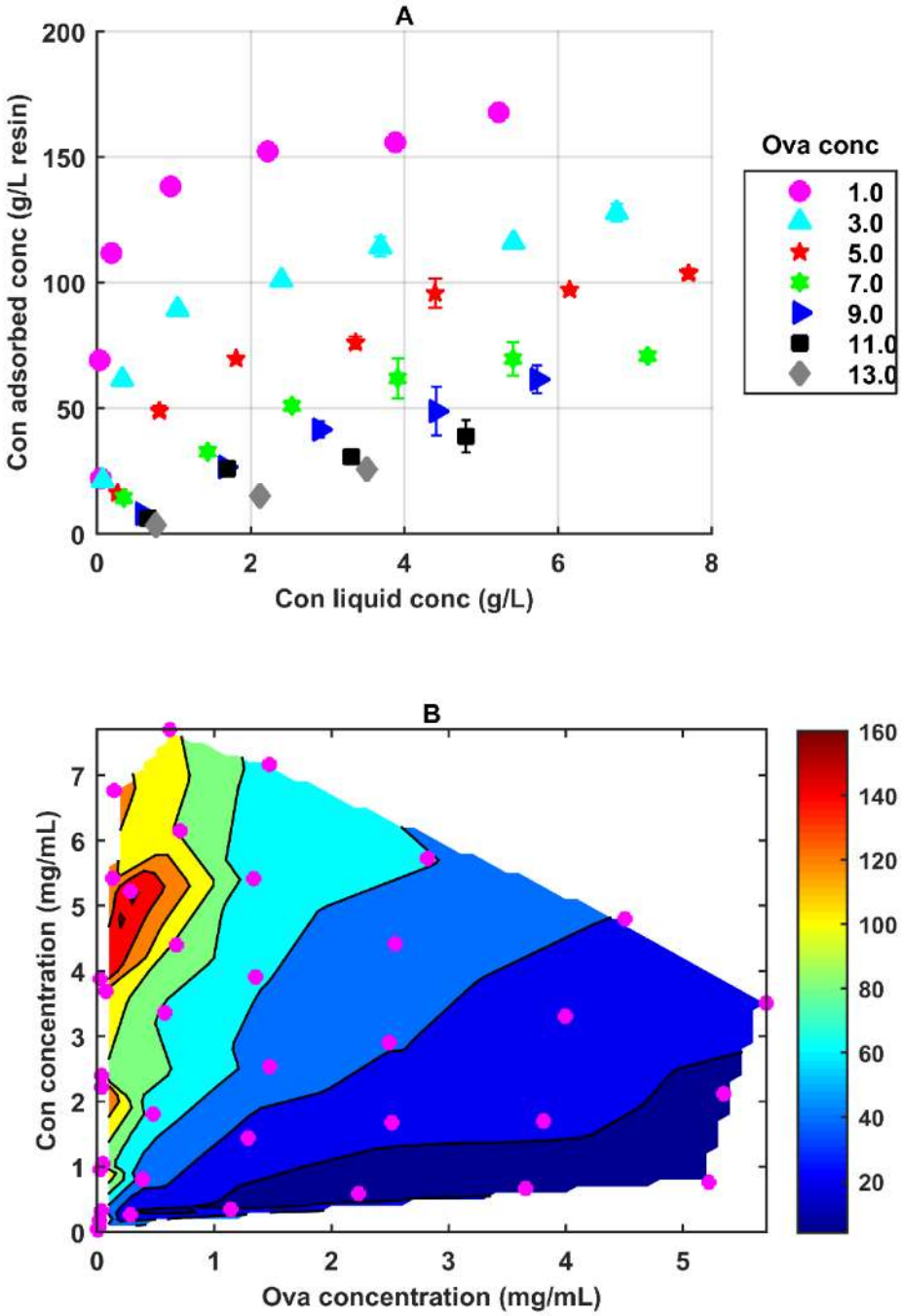

Figure 11-11 Con half of BSA-Con binary isotherms at 25 mM NaCl using q data only for adsorbed concentration estimate. A shows the isotherm plotted as if a series of single component isotherm wither different levels of start concentration of Ova represented in the legend in mg/mL. B shows the isotherm as a contour plot. Much of the systematic error in q was around 0 for Con so q* measure was taken not q.*

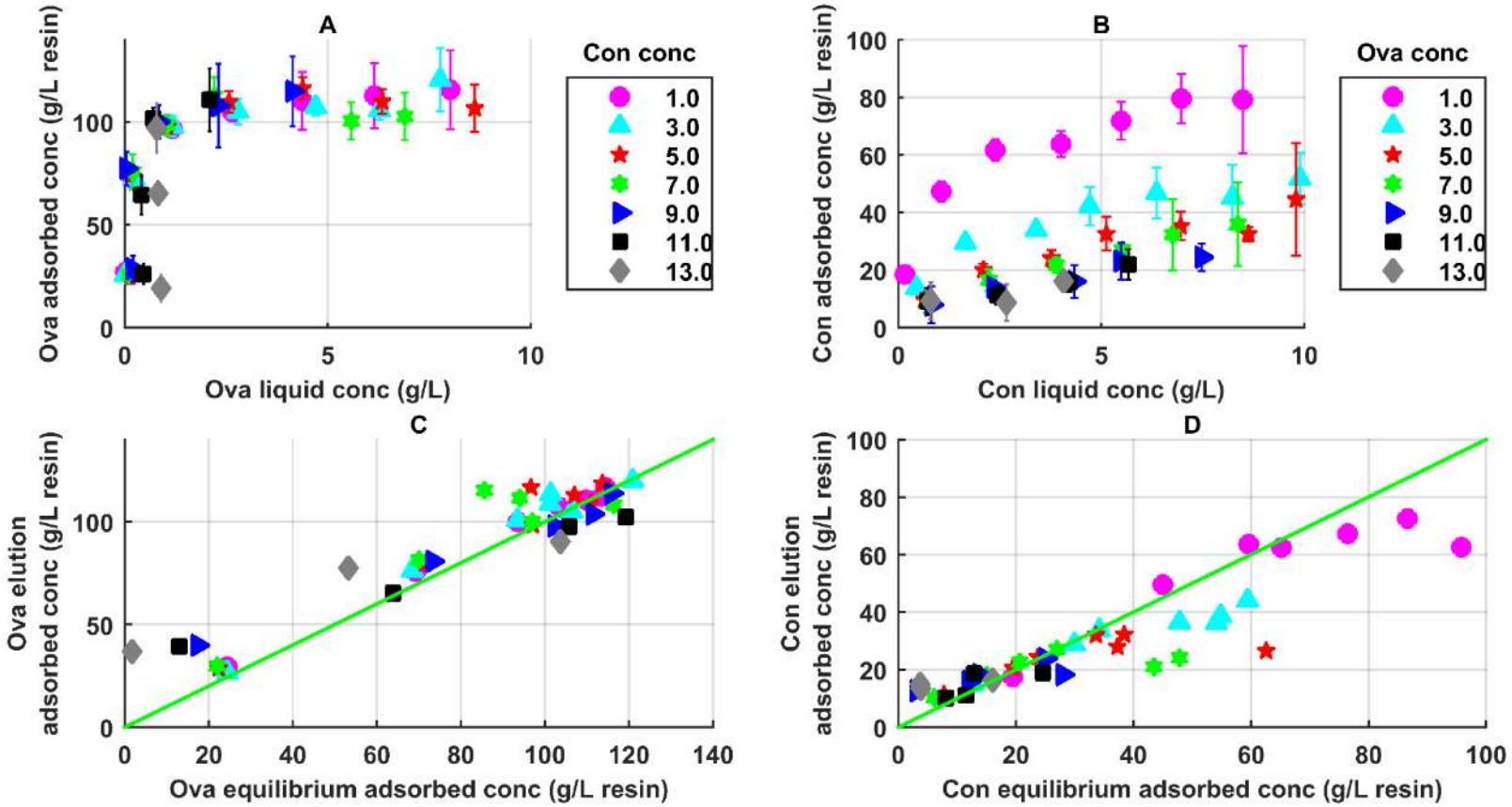

Figure 11-12 Ova-Con binary isotherm pH 9 50 mM NaCl. Panels A & C present Ova liquid and adsorbed concentrations, B & D represent Con data. Panels A and B present the binary isotherms at different starting concentrations of the competitor displayed in the figure legend in mg/mL, the liquid concentration is the average of q and q after outlier rejection. Error bars in panels A and B represent experimental standard deviation. Panels C and D present agreement between q and q*with a line of parity in green. Noise in the data set and agreement between q and q* is poor so both systematic and random error are displayed in Figure 11-13 to Figure 11-17.*

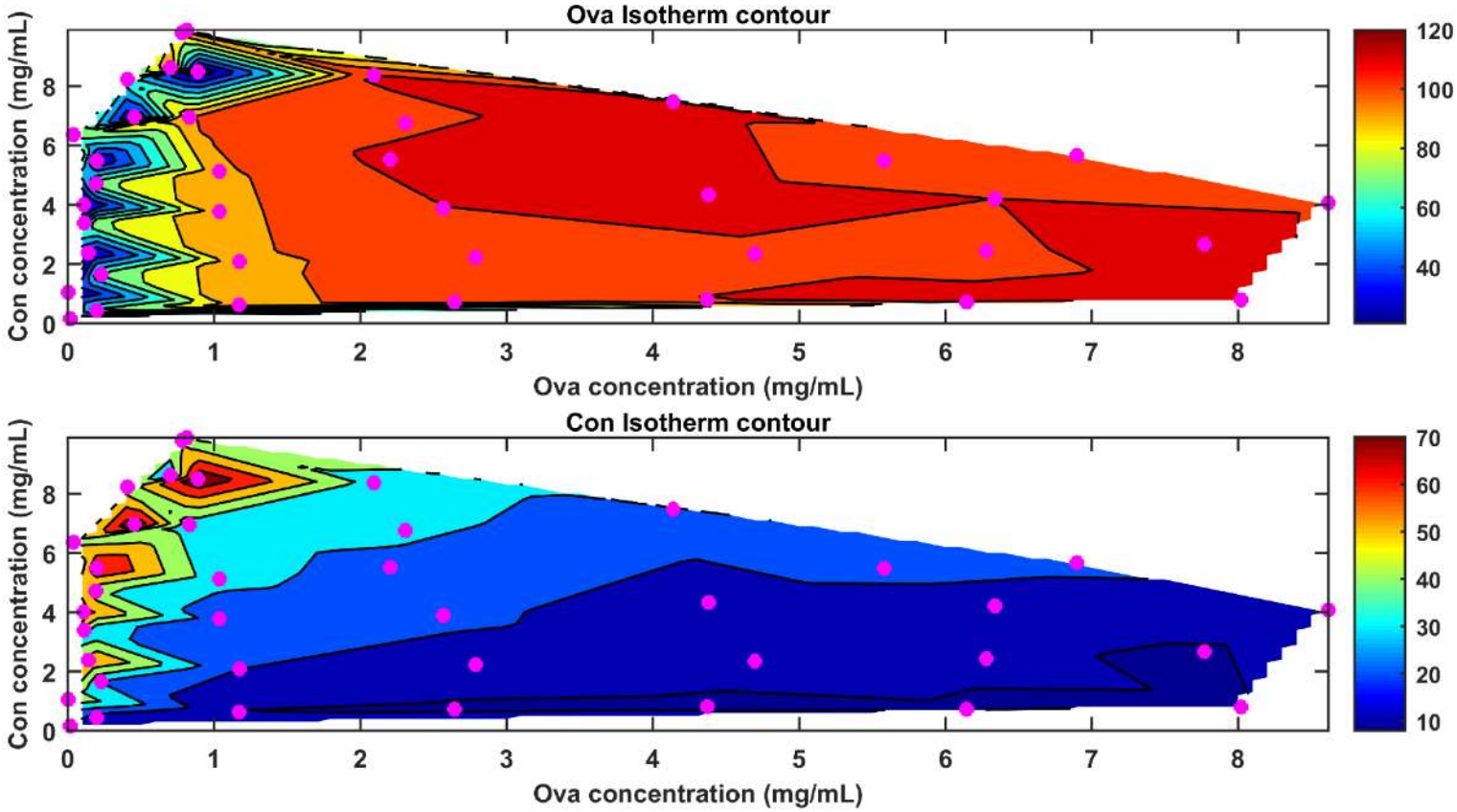

Figure 11-13 Binary isotherm for Ova-Con pH 9 50 mM NaCl. Isotherm has been drawn as a contour plot as a reference for error propagation plots. x and y axis display average equilibrium concentrations of Ova and Con measured directly after outlier rejection, contours display adsorbed concentration of Ova in the top panel and Con in the bottom panel, these have been taken as averages of q and q.*

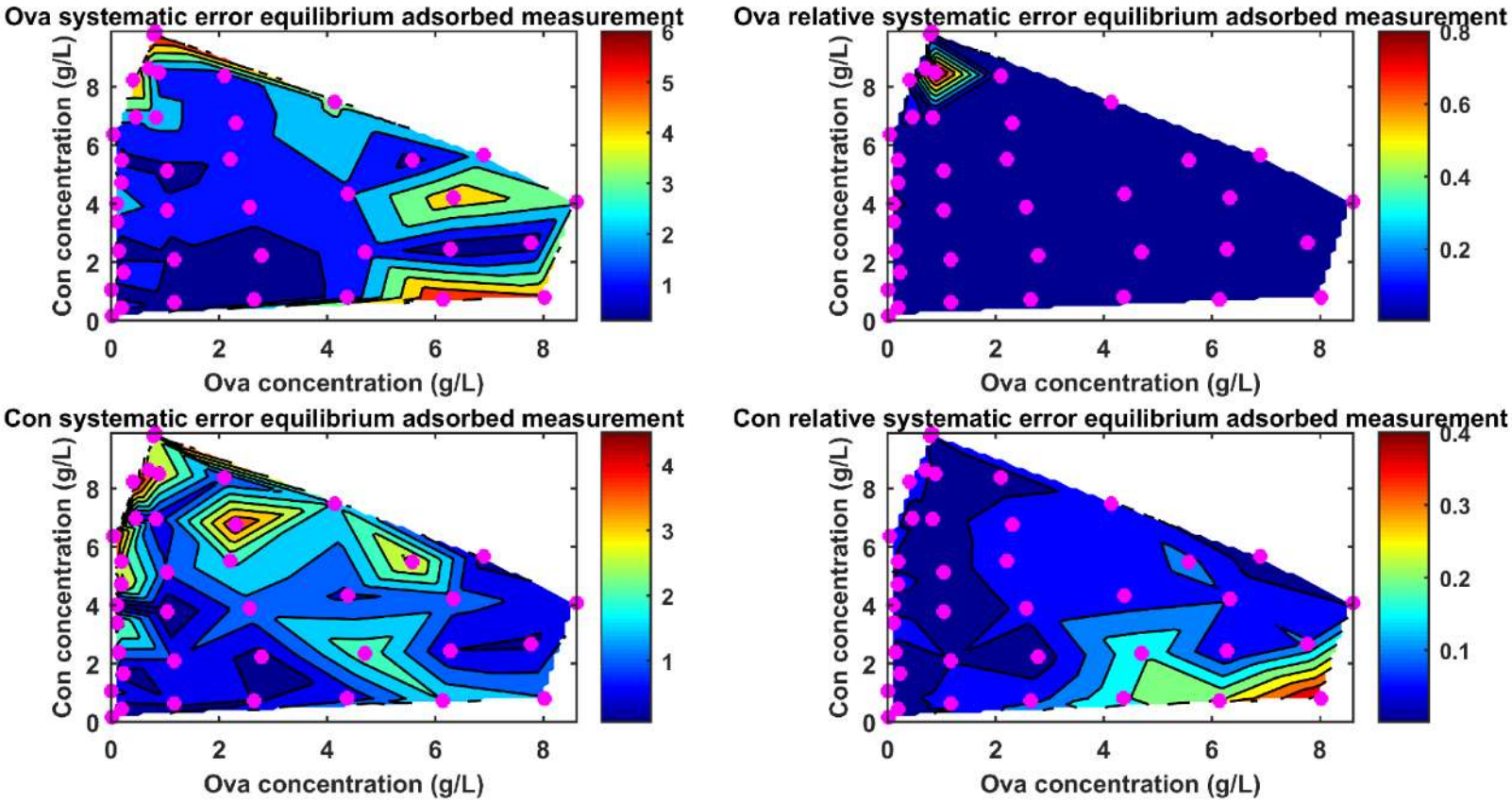

Figure 11-14 Contour plots showing propagated systematic absolute error in adsorbed concentration calculated using q for Ova-Con binary isotherms at pH 9 50 mM NaCl. Top 2 panels show propagated error for Ova and bottom 2 for Con. Left hand panels show propagated error in adsorbed concentration mg/mL resin and right hand panels show relative error in adsorbed concentration.

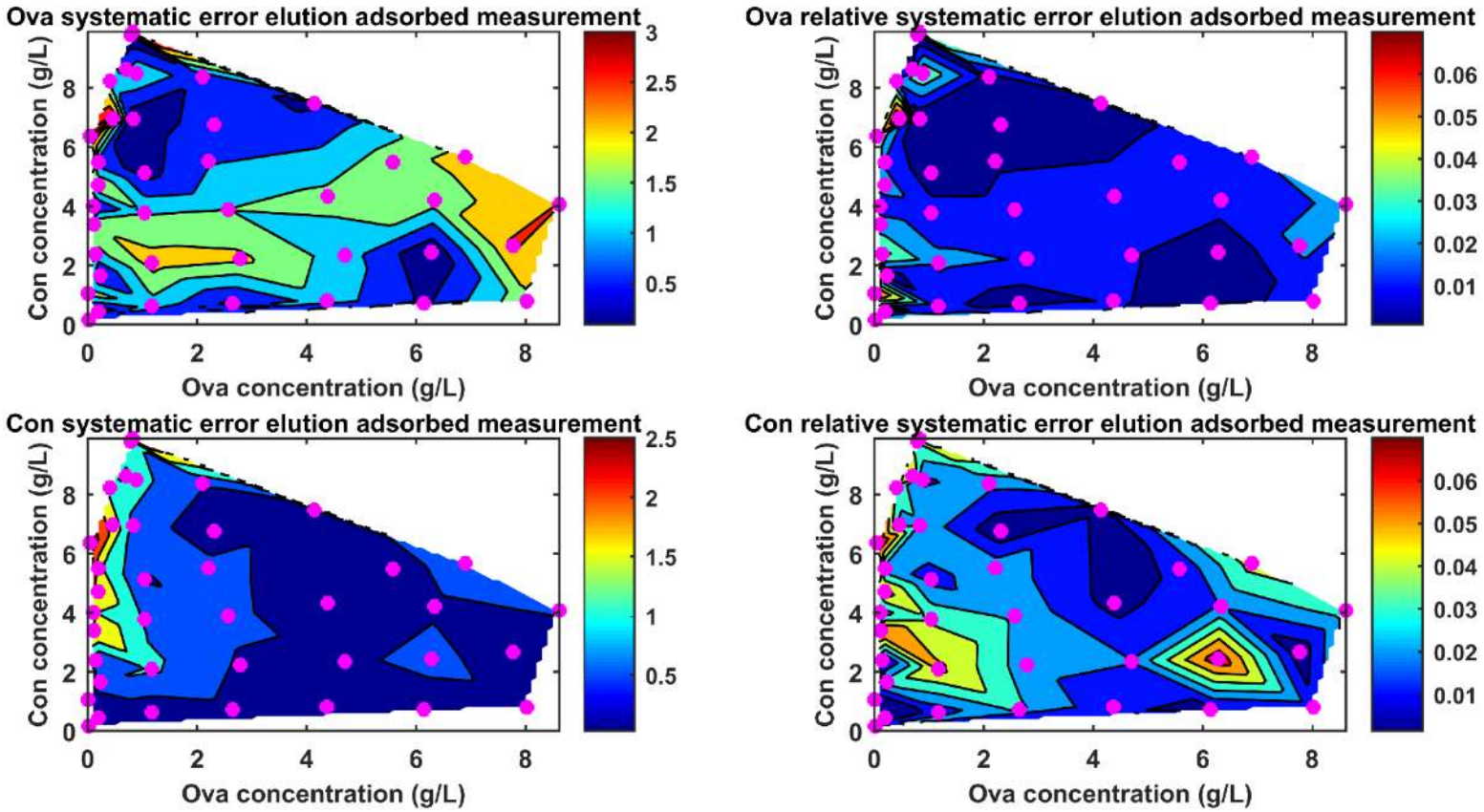

*Figure 11-15 Contour plots showing propagated systematic absolute error in adsorbed concentration calculated using q*for Ova-Con binary isotherms at pH 9 50 mM NaCl. Top 2 panels show propagated error for Ova and bottom 2 for Con. Left hand panels show propagated error in adsorbed concentration mg/mL resin and right hand panels show relative error in adsorbed concentration.*

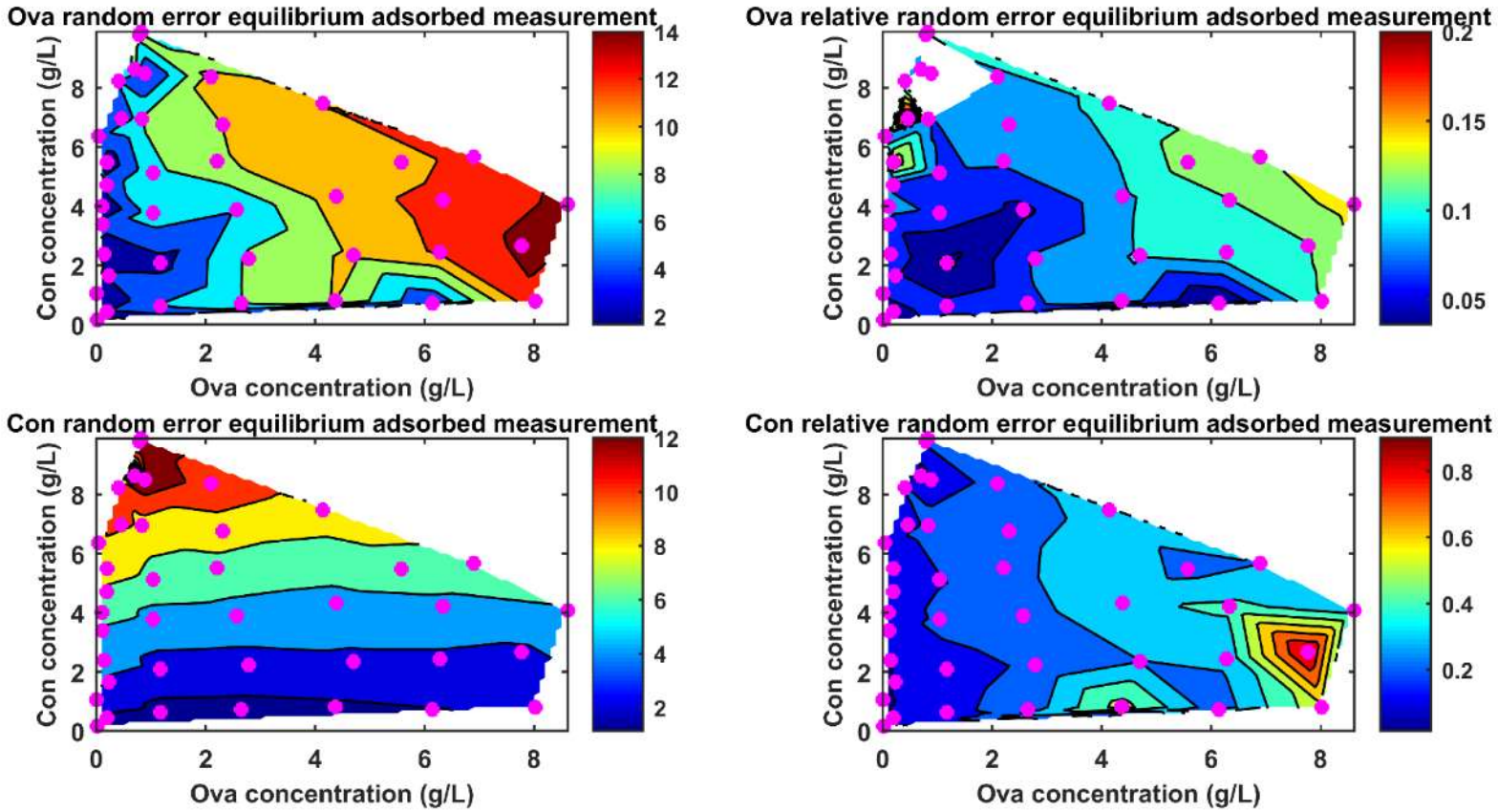

Figure 11-16 Contour plots showing propagated random error in adsorbed concentration calculated using q for Ova-Con binary isotherms at pH 9 50 mM NaCl. Top 2 panels show propagated error for Ova and bottom 2 for Con. Left hand panels show propagated error in adsorbed concentration mg/mL resin and right hand panels show relative error in adsorbed concentration.

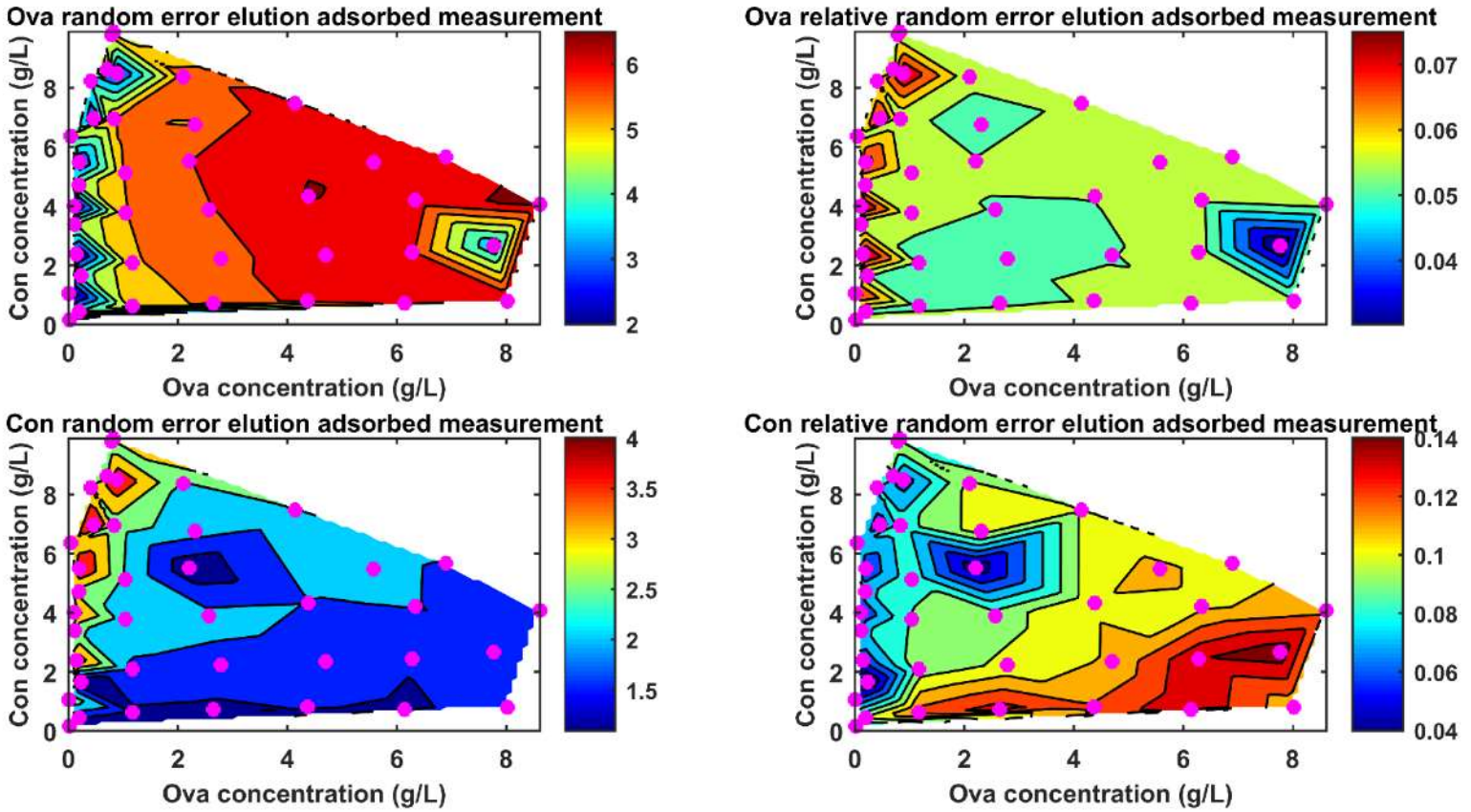

Figure 11-17 Contour plots showing propagated random error in adsorbed concentration calculated using q for Ova-Con binary isotherms at pH 9 50 mM NaCl. Top 2 panels show propagated error for Ova and bottom 2 for Con. Left hand panels show propagated error in adsorbed concentration mg/mL resin and right hand panels show relative error in adsorbed concentration.*

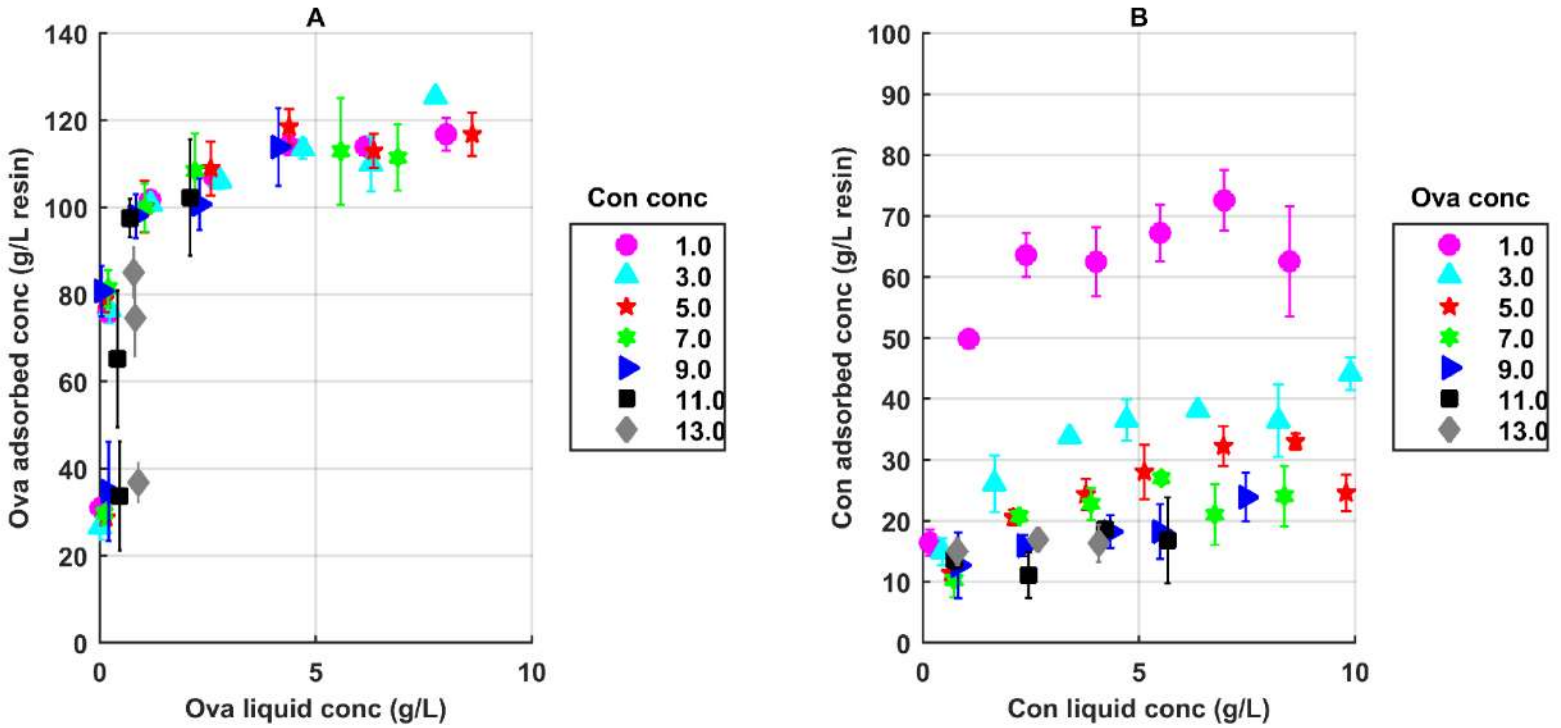

Figure 11-18 Ova-Con binary isotherm pH 9 50 mM NaCl. Panel A represents Ova data & B present Con liquid and adsorbed concentrations. Panels A and B present the binary isotherms at different starting concentrations of the competitor displayed in the figure legend in mg/mL, the liquid concentration is the assayed equilibrium average after outlier rejection, adsorbed concentration displays q data only. Error bars in panels A and B represent experimental standard deviation. Error propagation analysis suggests errors are greater for q measurement so q* data only is presented here.*

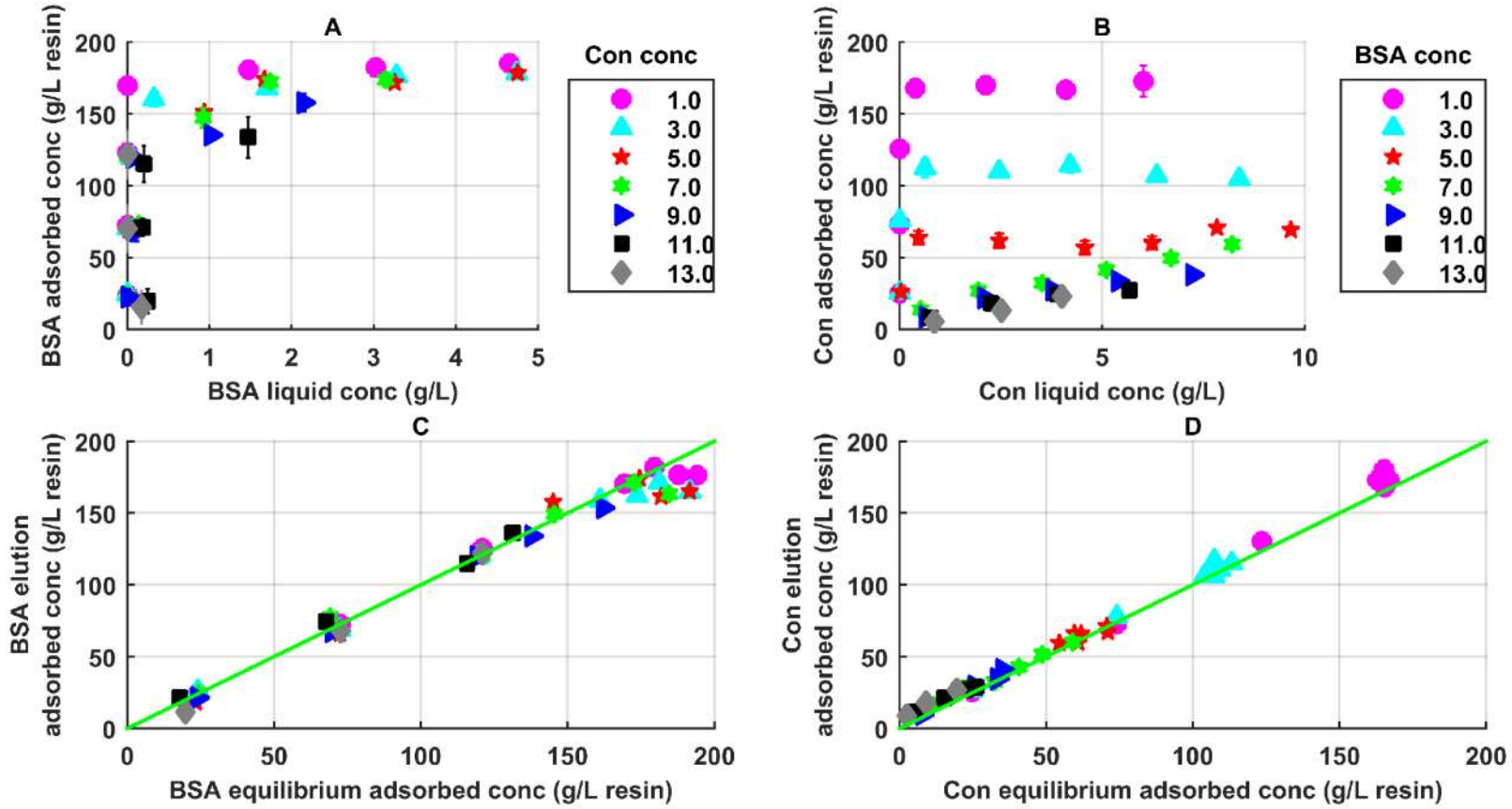

Figure 11-19 BSA-Con binary isotherm pH 9 0 mM NaCl. Panels A & C present BSA liquid and adsorbed concentrations, B & D represent Con data. Panels A and B present the binary isotherms at different starting concentrations of the competitor displayed in the figure legend in mg/mL, the liquid concentration is the average of q and q after outlier rejection. Error bars in panels A and B represent experimental standard deviation. Panels C and D present agreement between q and q*with a line of parity in green.*

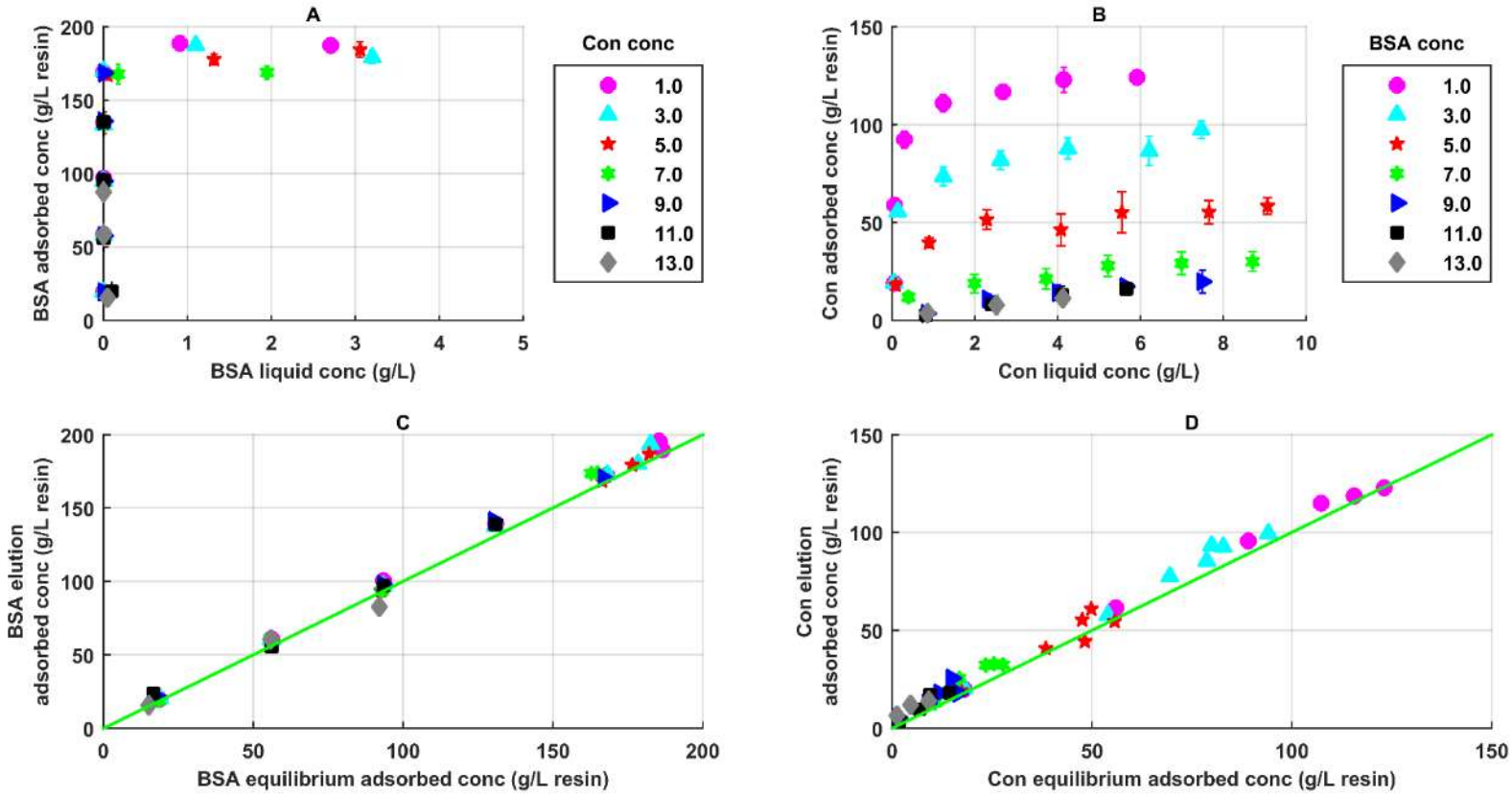

Figure 11-20 BSA-Con binary isotherm pH 9 25 mM NaCl. Panels A & C present BSA liquid and adsorbed concentrations, B & D represent Con data. Panels A and B present the binary isotherms at different starting concentrations of the competitor displayed in the figure legend in mg/mL, the liquid concentration is the average of q and q after outlier rejection. Error bars in panels A and B represent experimental standard deviation. Panels C and D present agreement between q and q*with a line of parity in green.*

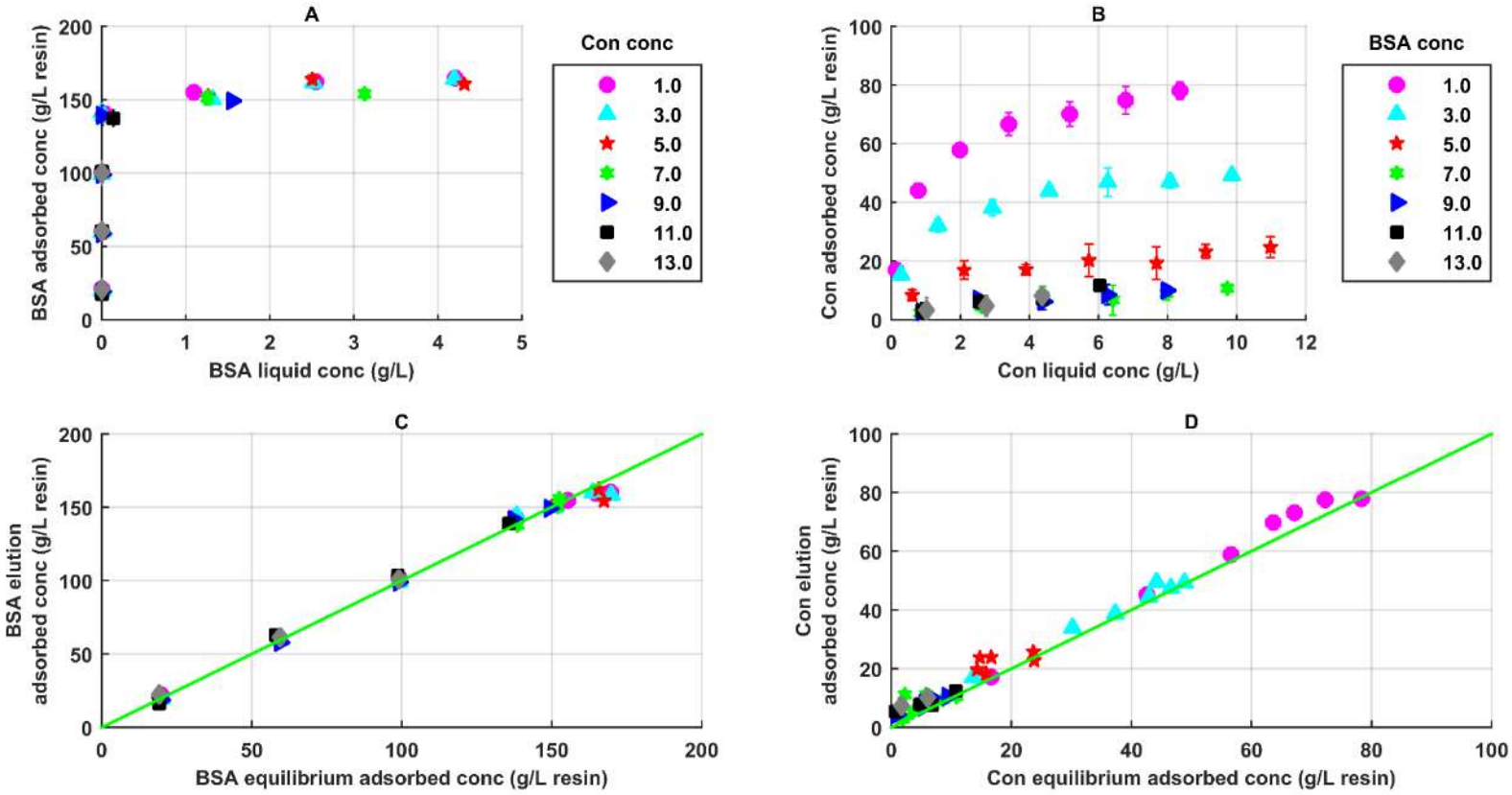

Figure 11-21 BSA-Con binary isotherm pH 9 50 mM NaCl. Panels A & C present BSA liquid and adsorbed concentrations, B & D represent Con data. Panels A and B present the binary isotherms at different starting concentrations of the competitor displayed in the figure legend in mg/mL, the liquid concentration is the average of q and q after outlier rejection. Error bars in panels A and B represent experimental standard deviation. Panels C and D present agreement between q and q*with a line of parity in green.*

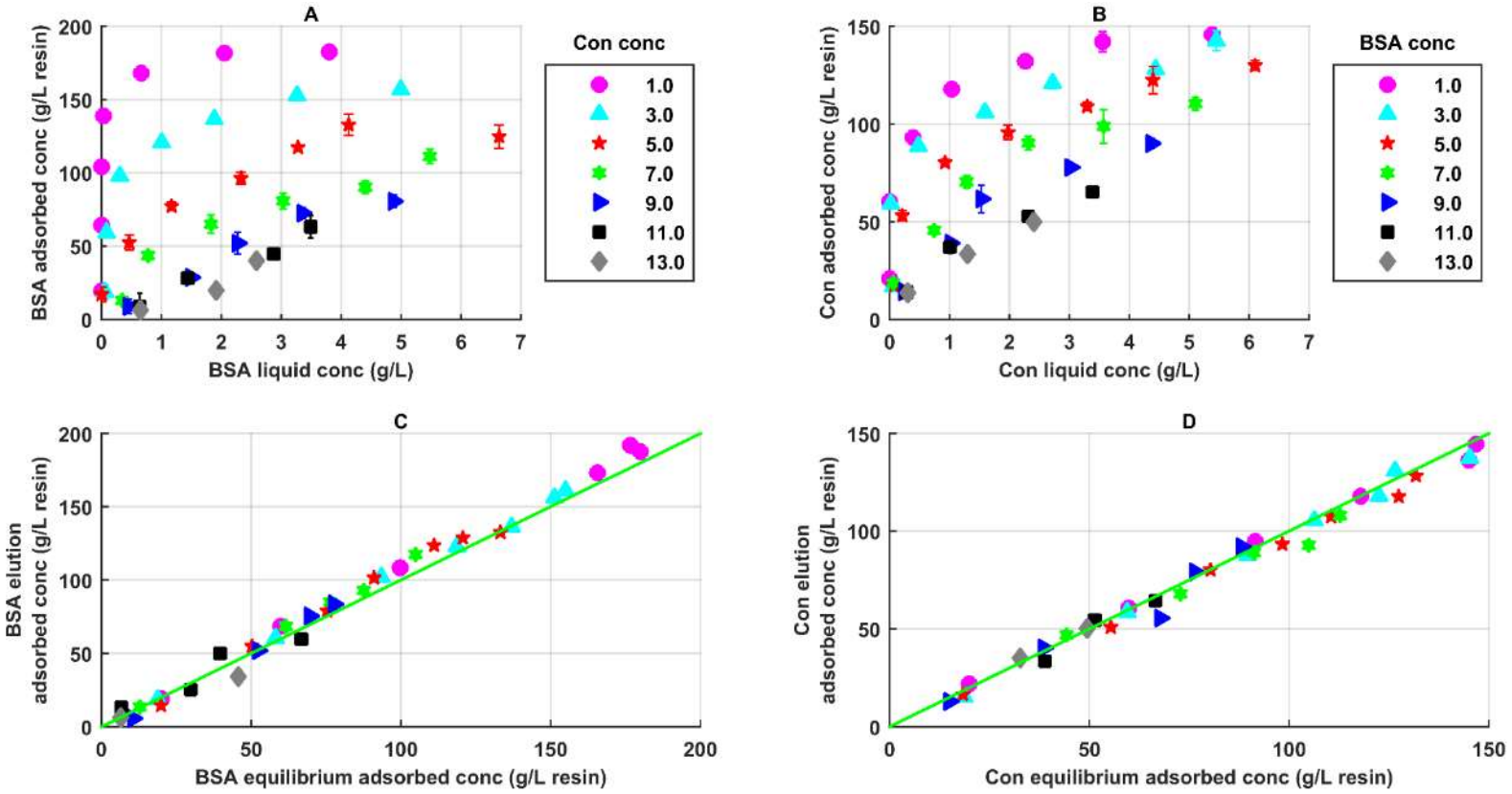

Figure 11-22 BSA-Ova binary isotherm pH 8 0 mM NaCl. Panels A & C present BSA liquid and adsorbed concentrations, B & D represent Ova data. Panels A and B present the binary isotherms at different starting concentrations of the competitor displayed in the figure legend in mg/mL, the liquid concentration is the average of q and q after outlier rejection. Error bars in panels A and B represent experimental standard deviation. Panels C and D present agreement between q and q*with a line of parity in green.*

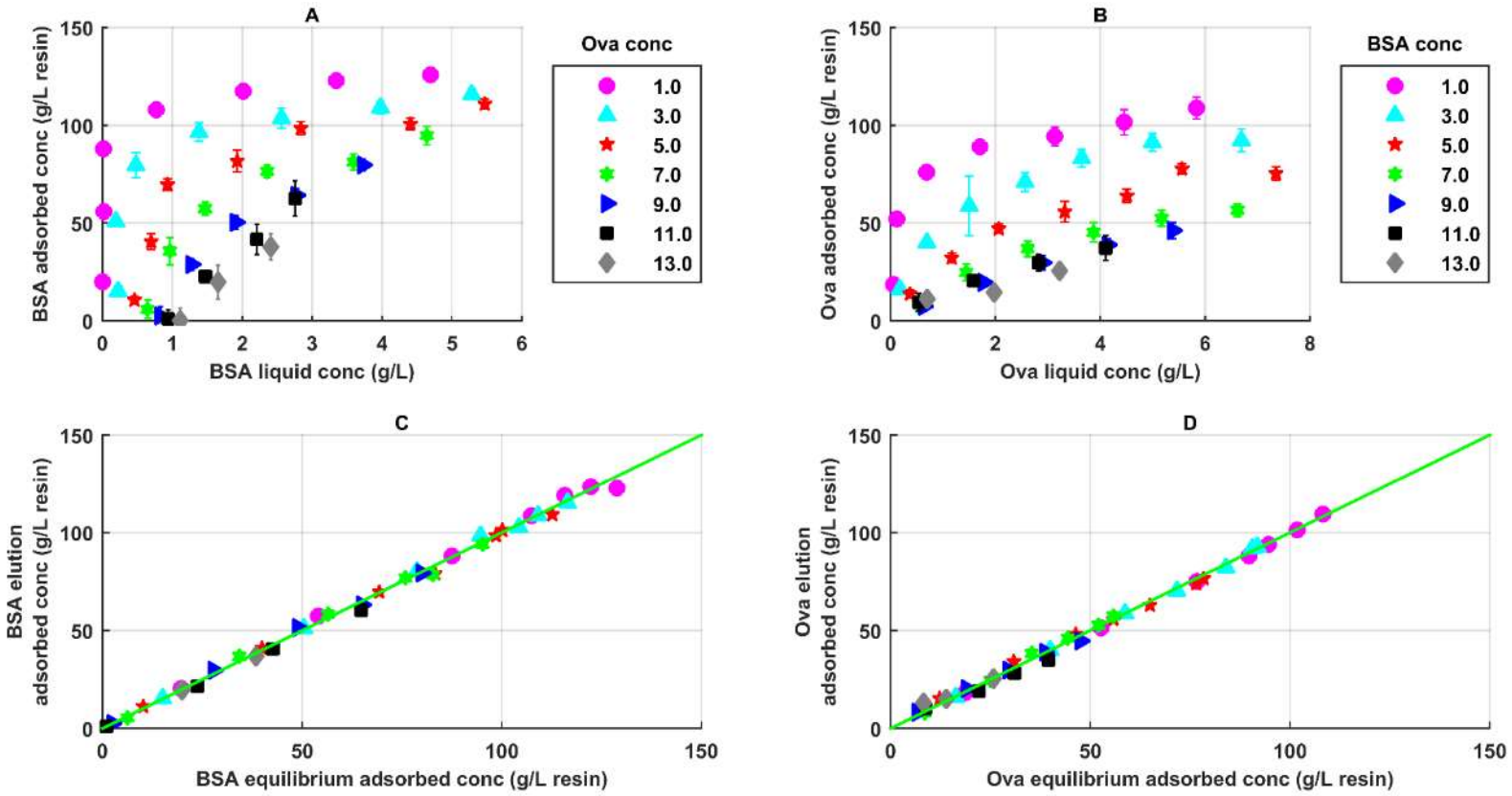

Figure 11-23 BSA-Ova binary isotherm pH 8 50 mM NaCl. Panels A & C present BSA liquid and adsorbed concentrations, B & D represent Ova data. Panels A and B present the binary isotherms at different starting concentrations of the competitor displayed in the figure legend in mg/mL, the liquid concentration is the average of q and q after outlier rejection. Error bars in panels A and B represent experimental standard deviation. Panels C and D present agreement between q and q*with a line of parity in green.*

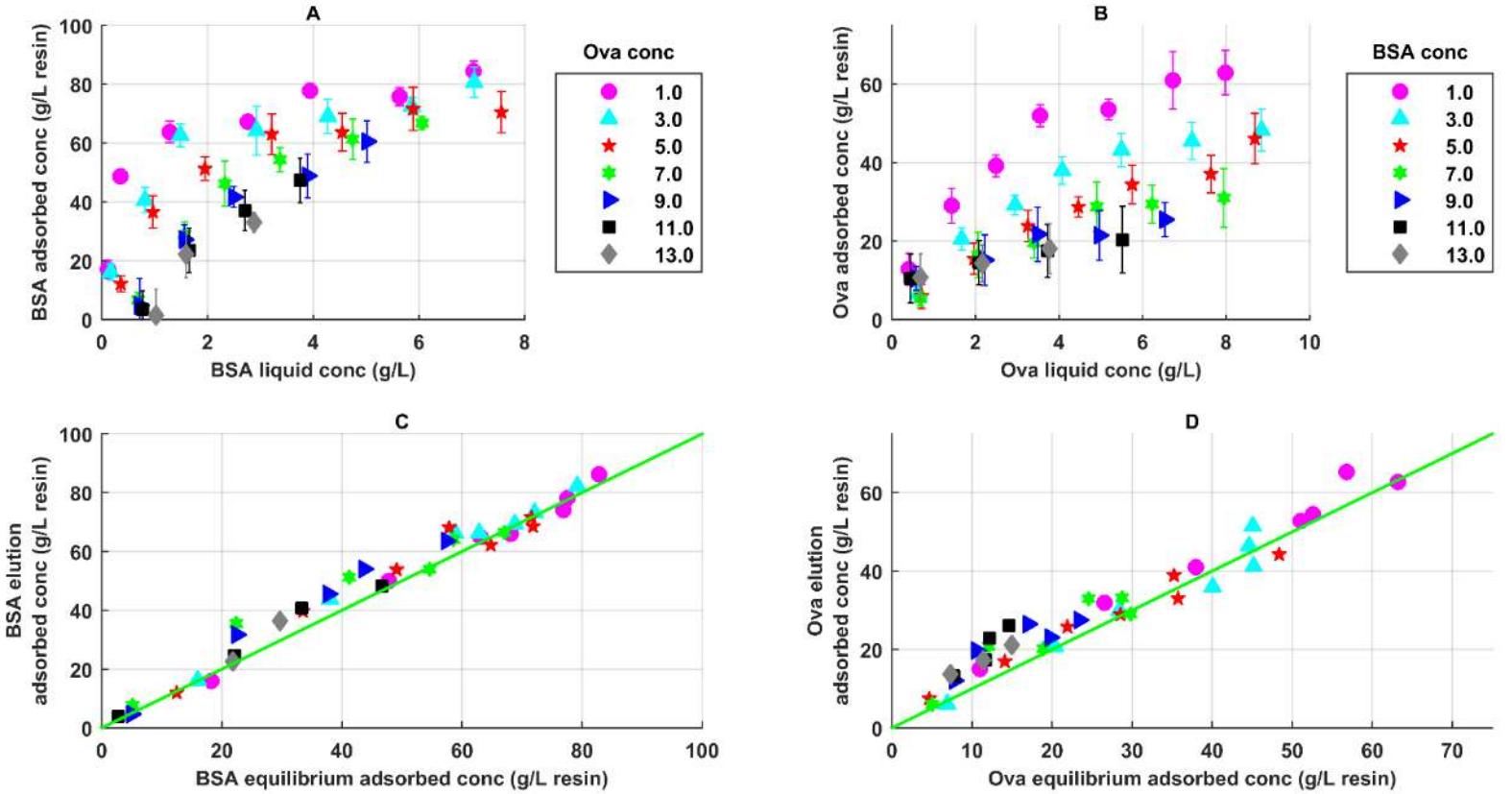

Figure 11-24 BSA-Ova binary isotherm pH 8 100 mM NaCl. Panels A & C present BSA liquid and adsorbed concentrations, B & D represent Ova data. Panels A and B present the binary isotherms at different starting concentrations of the competitor displayed in the figure legend in mg/mL, the liquid concentration is the average of q and q after outlier rejection. Error bars in panels A and B represent experimental standard deviation. Panels C and D present agreement between q and q*with a line of parity in green.*

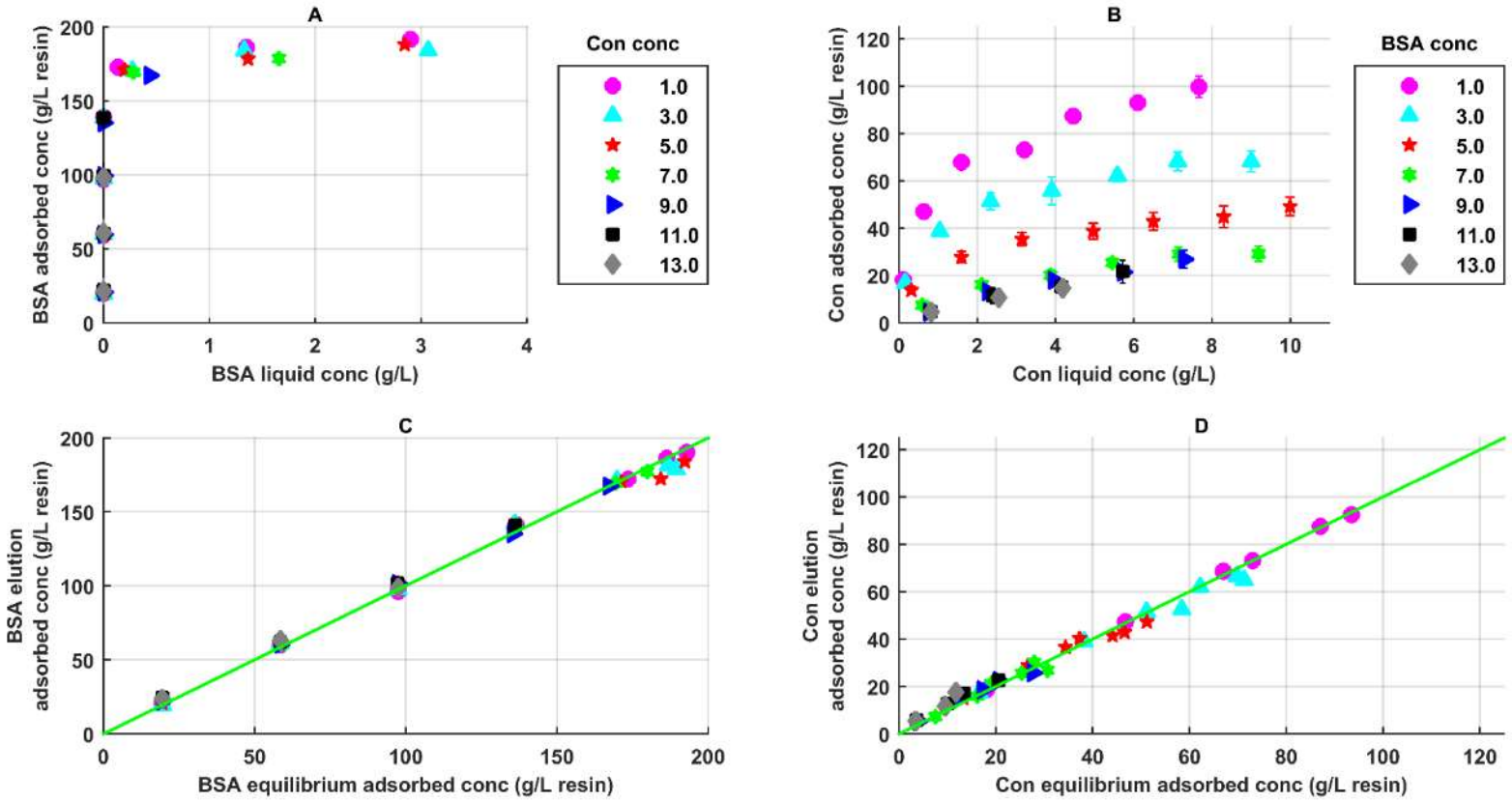

Figure 11-25 BSA-Con binary isotherm pH 8 0 mM NaCl. Panels A & C present BSA liquid and adsorbed concentrations, B & D represent Con data. Panels A nd B present the binary isotherms at different starting concentrations of the competitor displayed in the figure legend in mg/mL, the liquid concentration is the average of directly assayed concentrations after outlier rejection and the adsorbed concentration is the average of equilibrium and elution data after outlier rejection. Error bars in panels A and B represent experimental standard deviation. Panels C and D present agreement between adsorbed concentrations calculated at equilibrium and via elution with a line of parity in green.

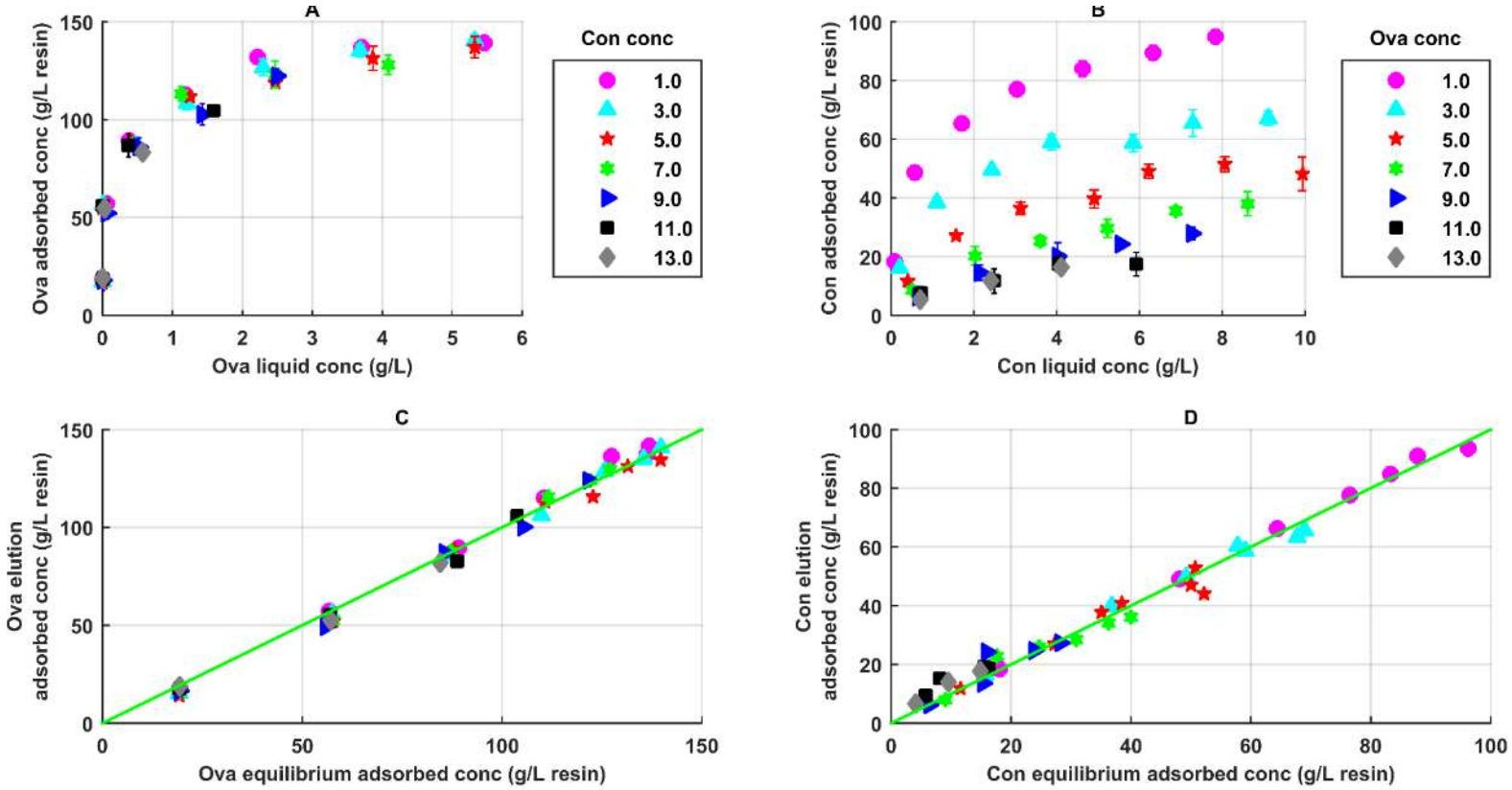

Figure 11-26 Ova-Con binary isotherm pH 8 0 mM NaCl. Panels A & C present Ova liquid and adsorbed concentrations, B & D represent Con data. Panels A nd B present the binary isotherms at different starting concentrations of the competitor displayed in the figure legend in mg/mL, the liquid concentration is the average of directly assayed concentrations after outlier rejection and the adsorbed concentration is the average of equilibrium and elution data after outlier rejection. Error bars in panels A and B represent experimental standard deviation. Panels C and D present agreement between adsorbed concentrations calculated at equilibrium and via elution with a line of parity in green.

11.2 Appendix B: Additional ternary isotherm datasets

Ternary isotherms of BSA-Ova-Con were studied at pH 9 50 mM Tris and 0 mM 25 and 50 mM NaCl on Capto Q strong anion exchanger. Isotherm data is presented here is at 25 and 50 mM NaCl levels. The propagation of systematic and random errors are presented when the data suggests it is pertinent. Additionally, the datasets have also been trimmed and manual omission of points has been applied when necessary as described in Chapter 7 Ternary isotherms and the propagation of their error.

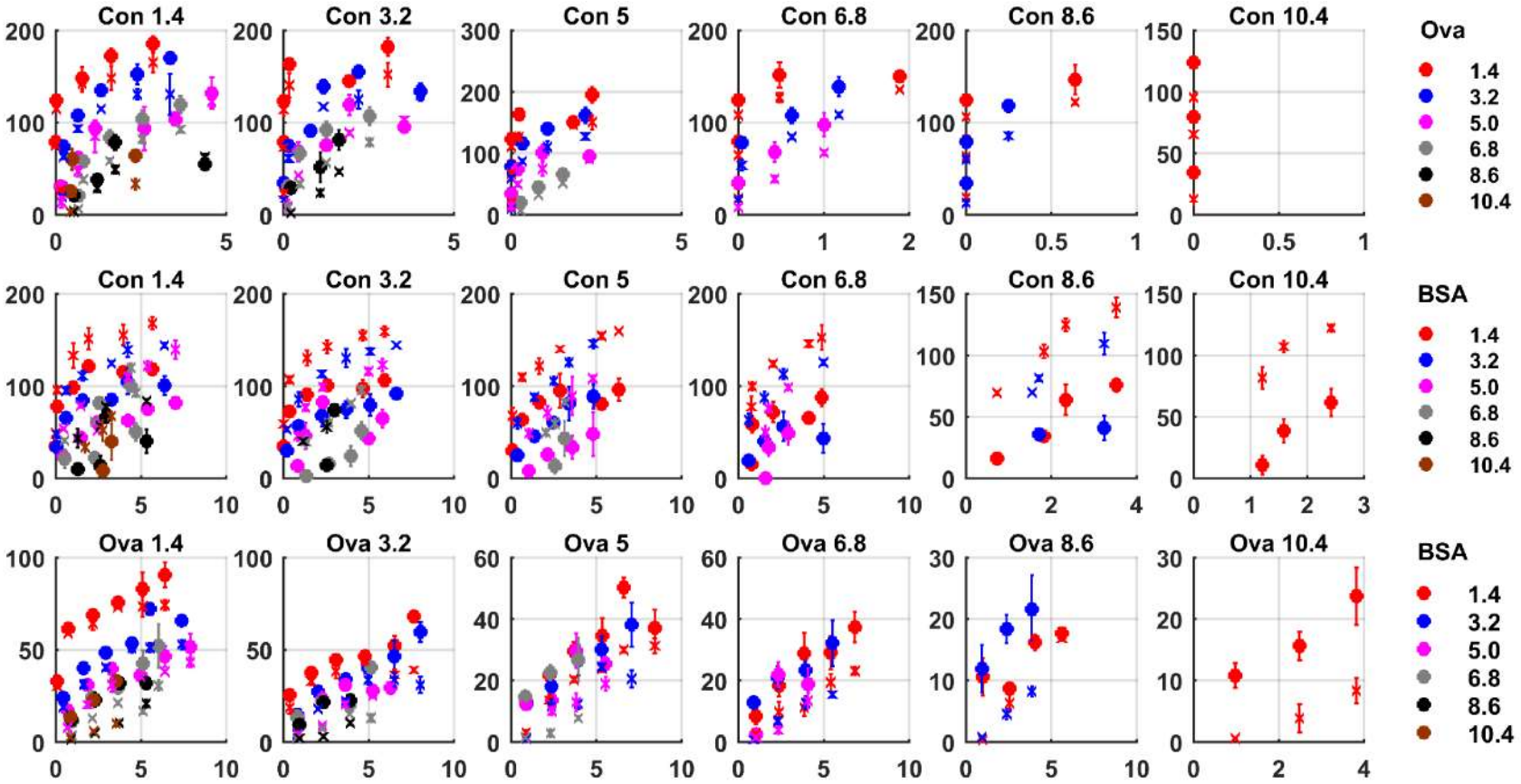

Figure 11-27 Ternary isotherm of BSA, Ova and Con plot as a series of single component isotherms at different starting concentrations of competitor protein at pH 9 and 25 mM NaCl. The plot titles display the starting concentration of one of the competitors and the legends display the starting concentration of the second competing protein in mg/mL. All x axis display liquid concentrations in mg/mL and all y-axis represent adsorbed concentration in mg/mL of resin. Top row of 6 plots display BSA isotherms, second row of 6 plots displays Ova isotherms and the third row of 6 plots displays Con isotherms. Circles represent q data and x represents q. Error bars represent standard deviation of experimental data. Poor agreement between q and q* is observed so systematic error propagation is included in Figure 11-28 to Figure 11-36.*

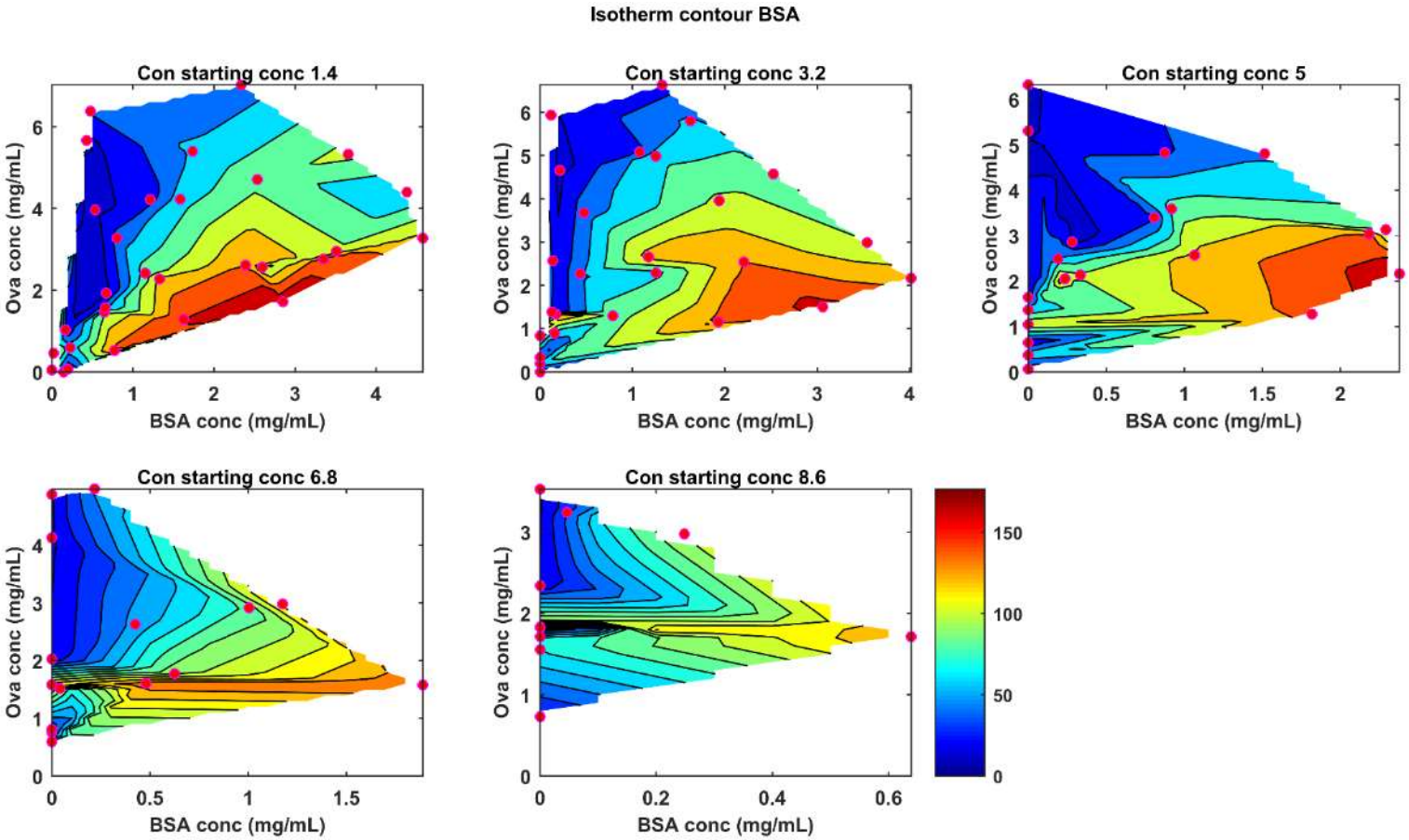

Figure 11-28 Ternary isotherm at pH 9 25 mM NaCl, contours display adsorbed concentration of BSA, adsorbed concentrations is as an average q and q. Isotherm has been drawn as a contour plot as a reference for error propagation plots. Each contour sub-plot can be thought of as a BSA-Ova binary isotherm at different starting concentrations of Con, the starting concentration of Con in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of BSA and Ova measured directly after outlier rejection.*

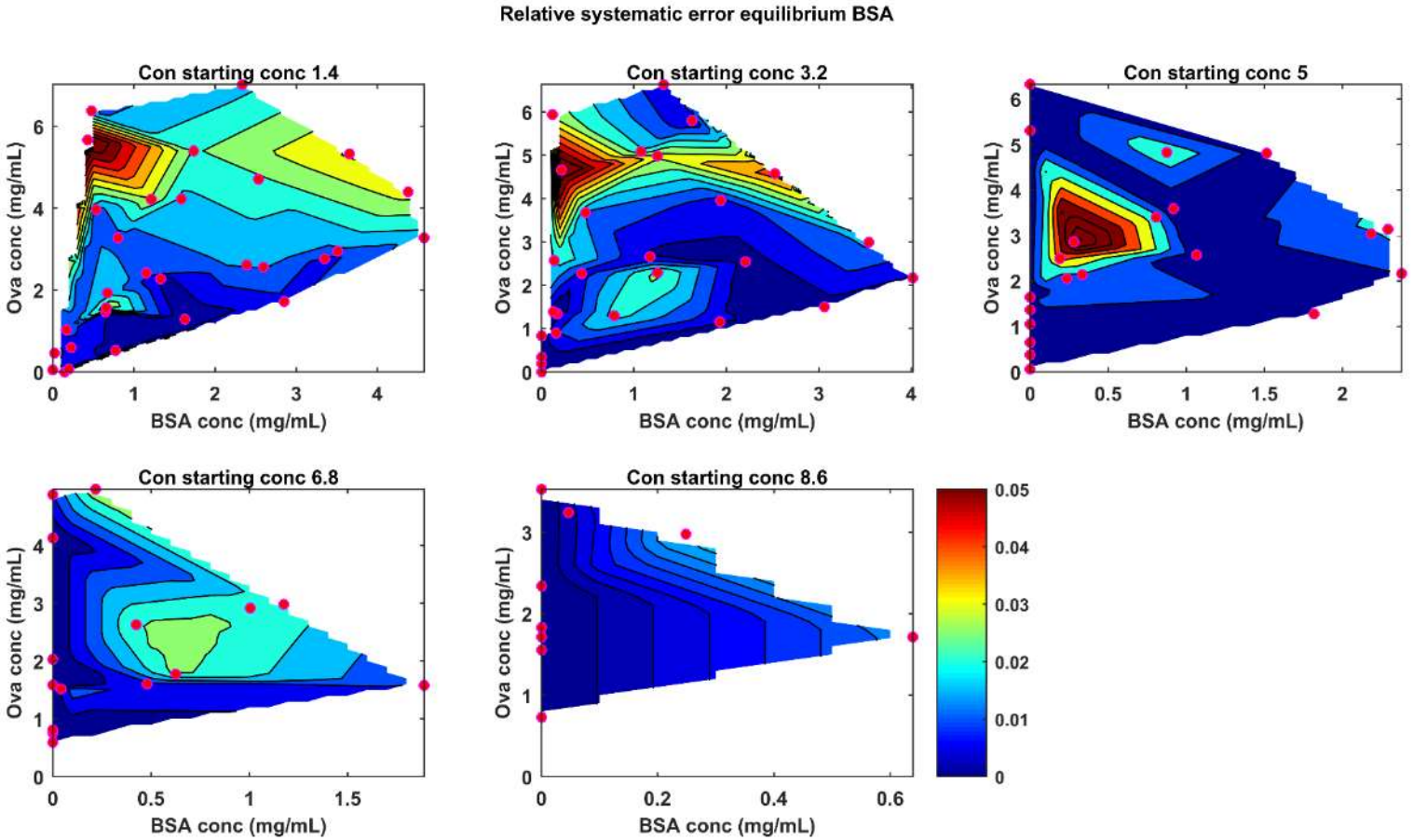

Figure 11-29 Contour plots showing propagated relative systematic error in q for BSA in ternary isotherms at pH 9 25 mM NaCl. The colours represent relative adsorbed error. Each contour sub-plot can be thought of as a BSA-Ova binary isotherm at different starting concentrations of Con, the starting concentration of Con in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of BSA and Ova measured directly after outlier rejection.

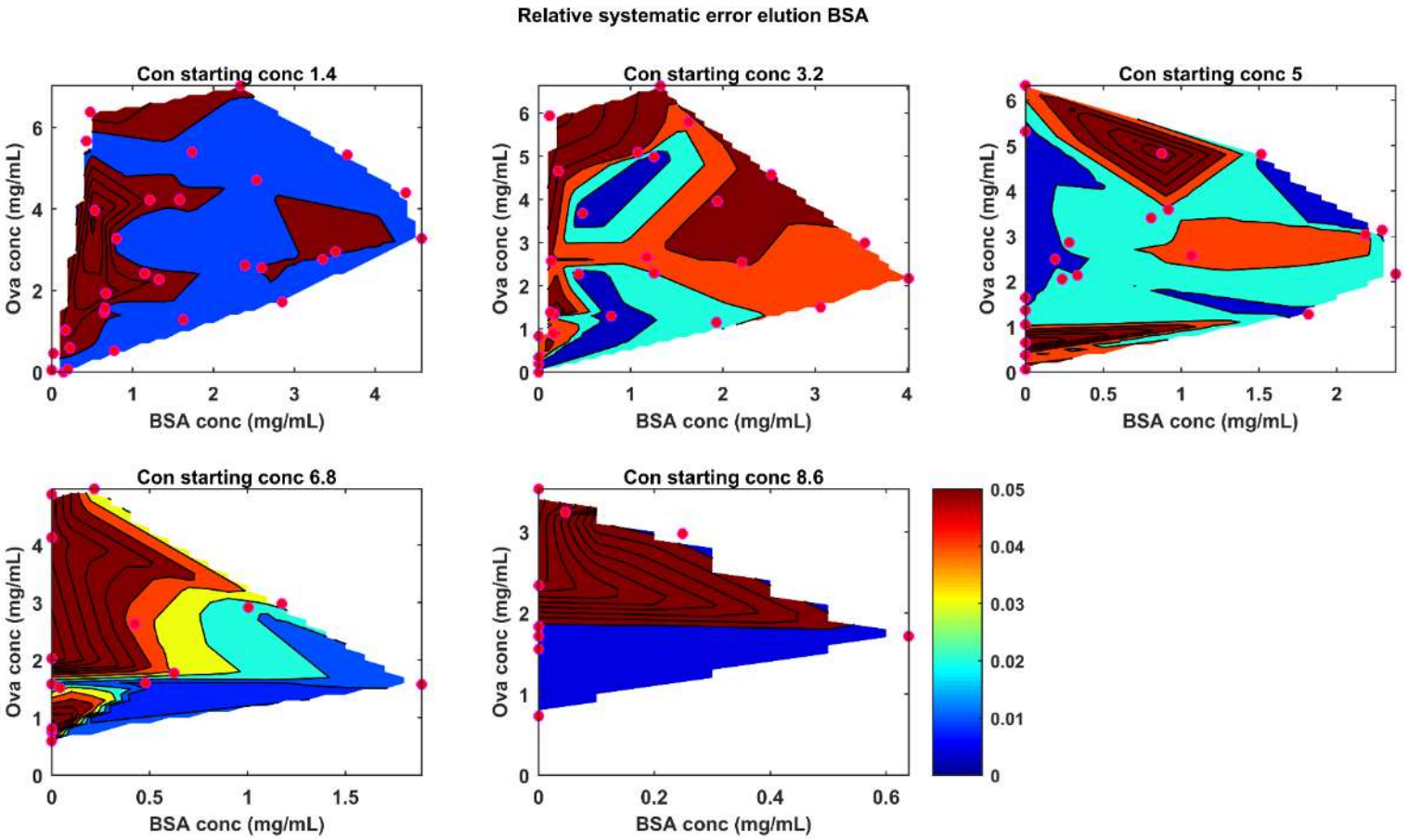

Figure 11-30 Contour plots showing propagated relative systematic error in q for BSA in ternary isotherms at pH 9 25 mM NaCl. The colours represent relative adsorbed error. Each contour sub-plot can be thought of as a BSA-Ova binary isotherm at different starting concentrations of Con, the starting concentration of Con in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of BSA and Ova measured directly after outlier rejection.*

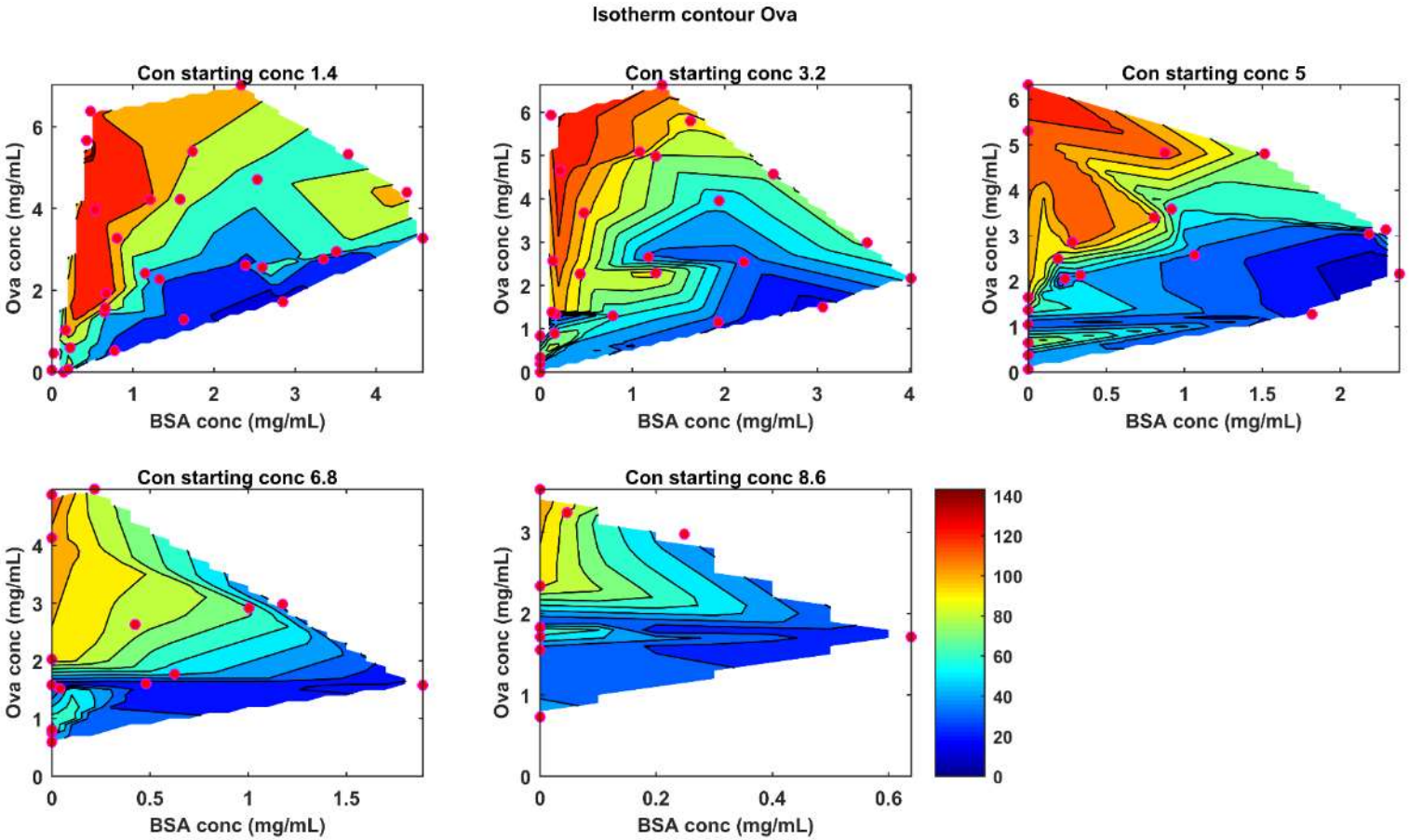

Figure 11-31 Ternary isotherm at pH 9 25 mM NaCl, contours display adsorbed concentration of Ova, adsorbed concentrations is as an average q and q. Isotherm has been drawn as a contour plot as a reference for error propagation plots. Each contour sub-plot can be thought of as a BSA-Ova binary isotherm at different starting concentrations of Con, the starting concentration of Con in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of BSA and Ova measured directly after outlier rejection.*

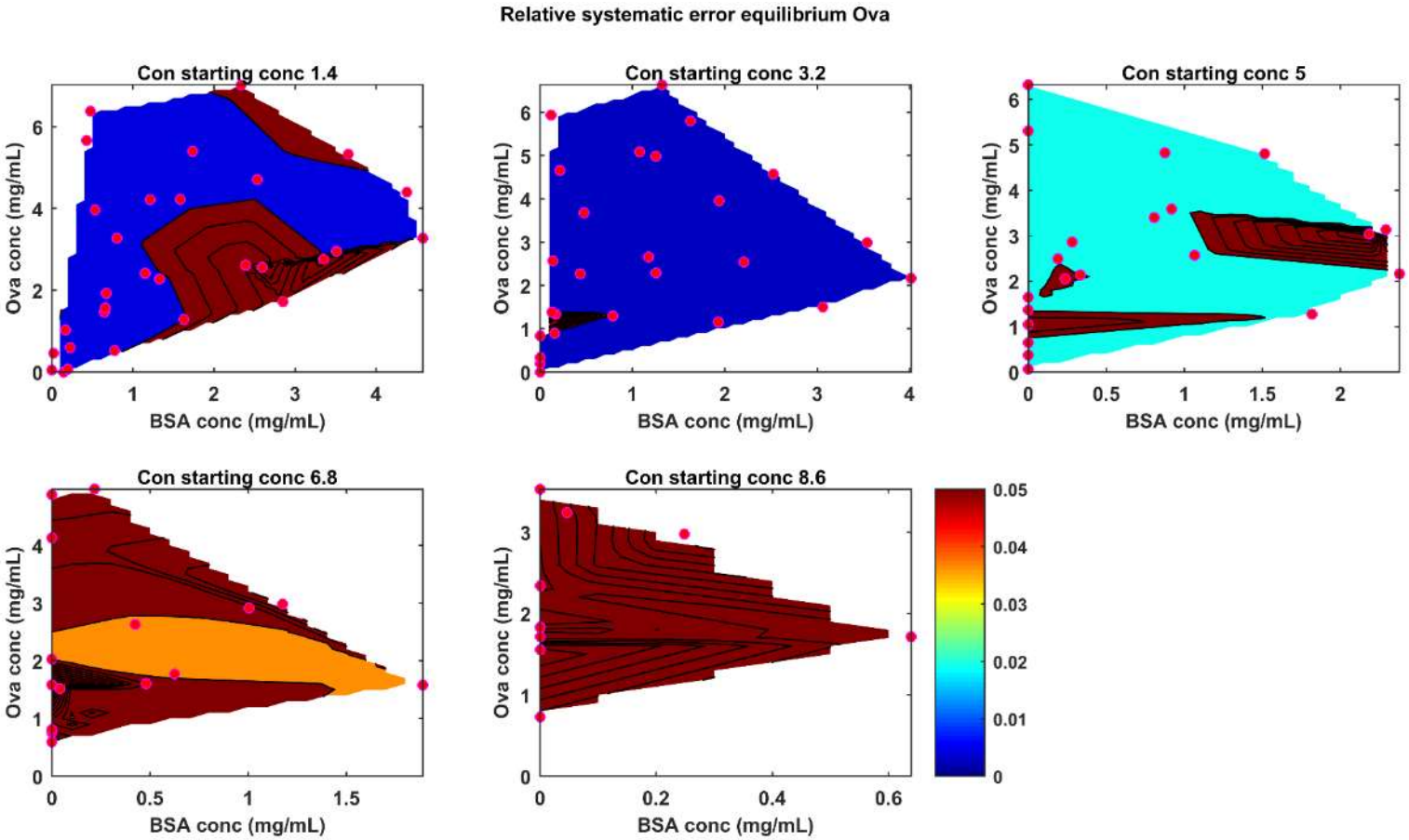

Figure 11-32 Contour plots showing propagated relative systematic error in q for Ova in ternary isotherms at pH 9 25 mM NaCl. The colours represent relative adsorbed error. Each contour sub-plot can be thought of as a BSA-Ova binary isotherm at different starting concentrations of Con, the starting concentration of Con in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of BSA and Ova measured directly after outlier rejection.

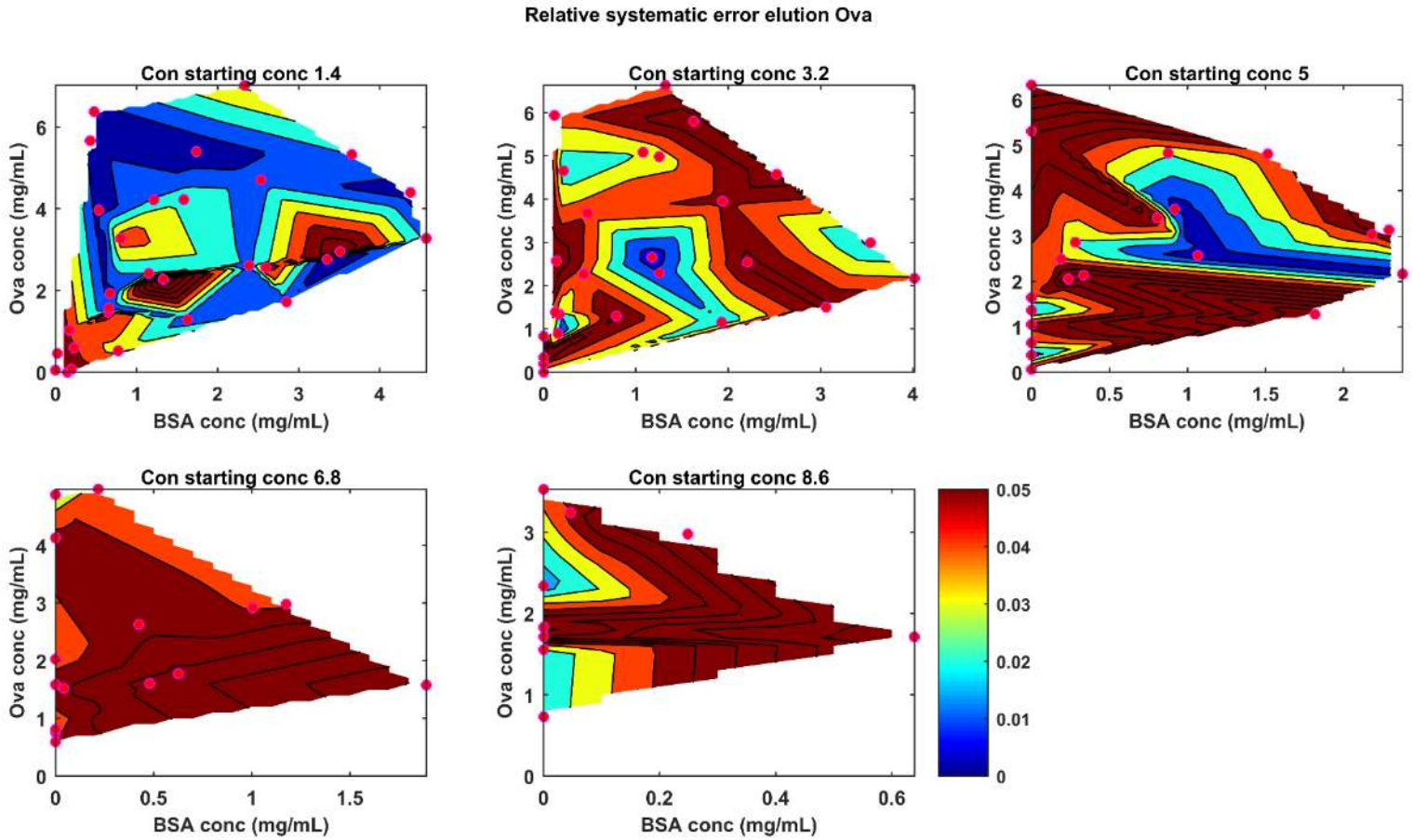

Figure 11-33 Contour plots showing propagated relative systematic error in q for Ova in ternary isotherms at pH 9 25 mM NaCl. The colours represent relative adsorbed error. Each contour sub-plot can be thought of as a BSA-Ova binary isotherm at different starting concentrations of Con, the starting concentration of Con in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of BSA and Ova measured directly after outlier rejection.*

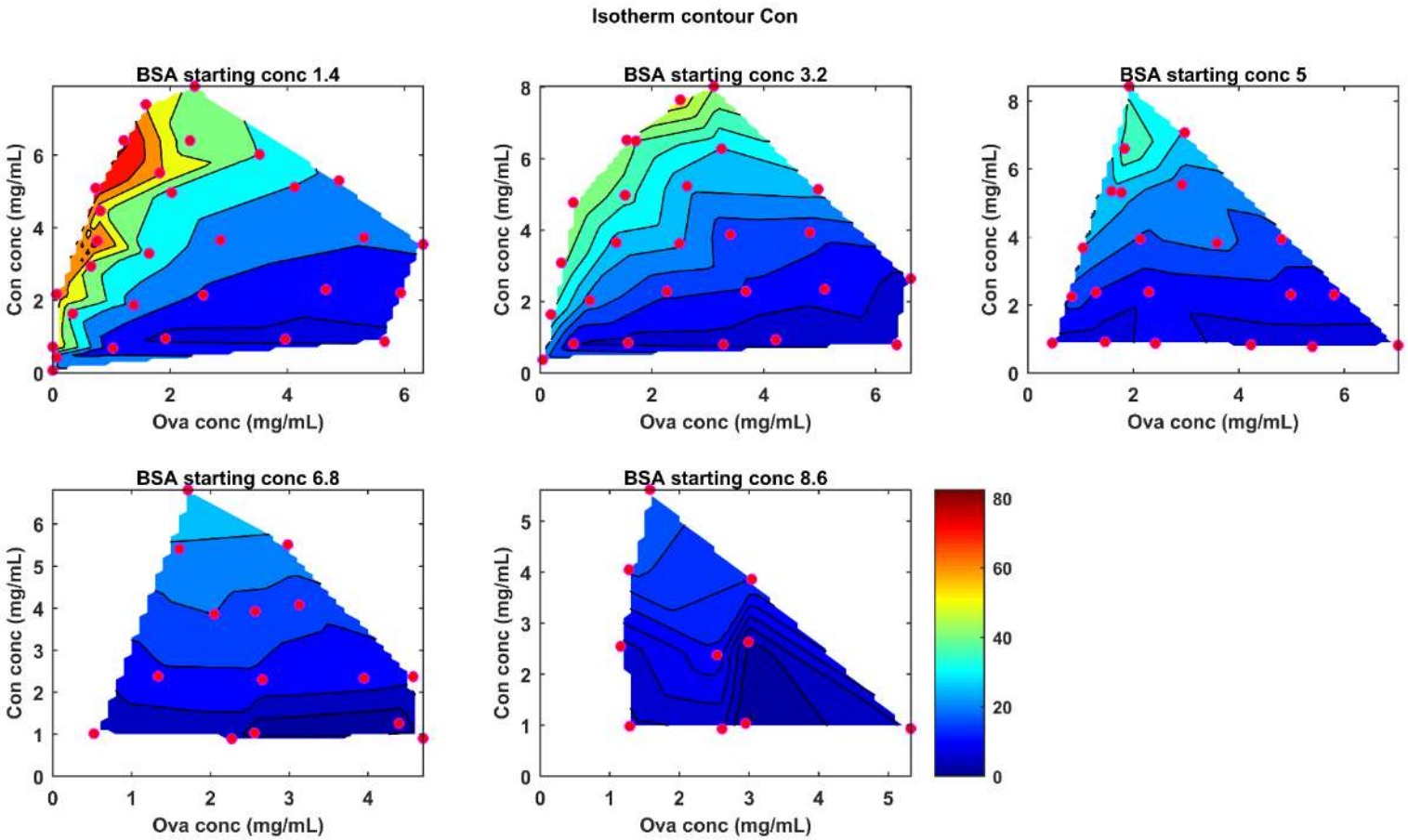

Figure 11-34 Ternary isotherm at pH 9 25 mM NaCl, contours display adsorbed concentration of Con, adsorbed concentrations is as an average q and q. Isotherm has been drawn as a contour plot as a reference for error propagation plots. Each contour sub-plot can be thought of as a Ova-Con binary isotherm at different starting concentrations of BSA, the starting concentration of BSA in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of Ova and Con measured directly after outlier rejection.*

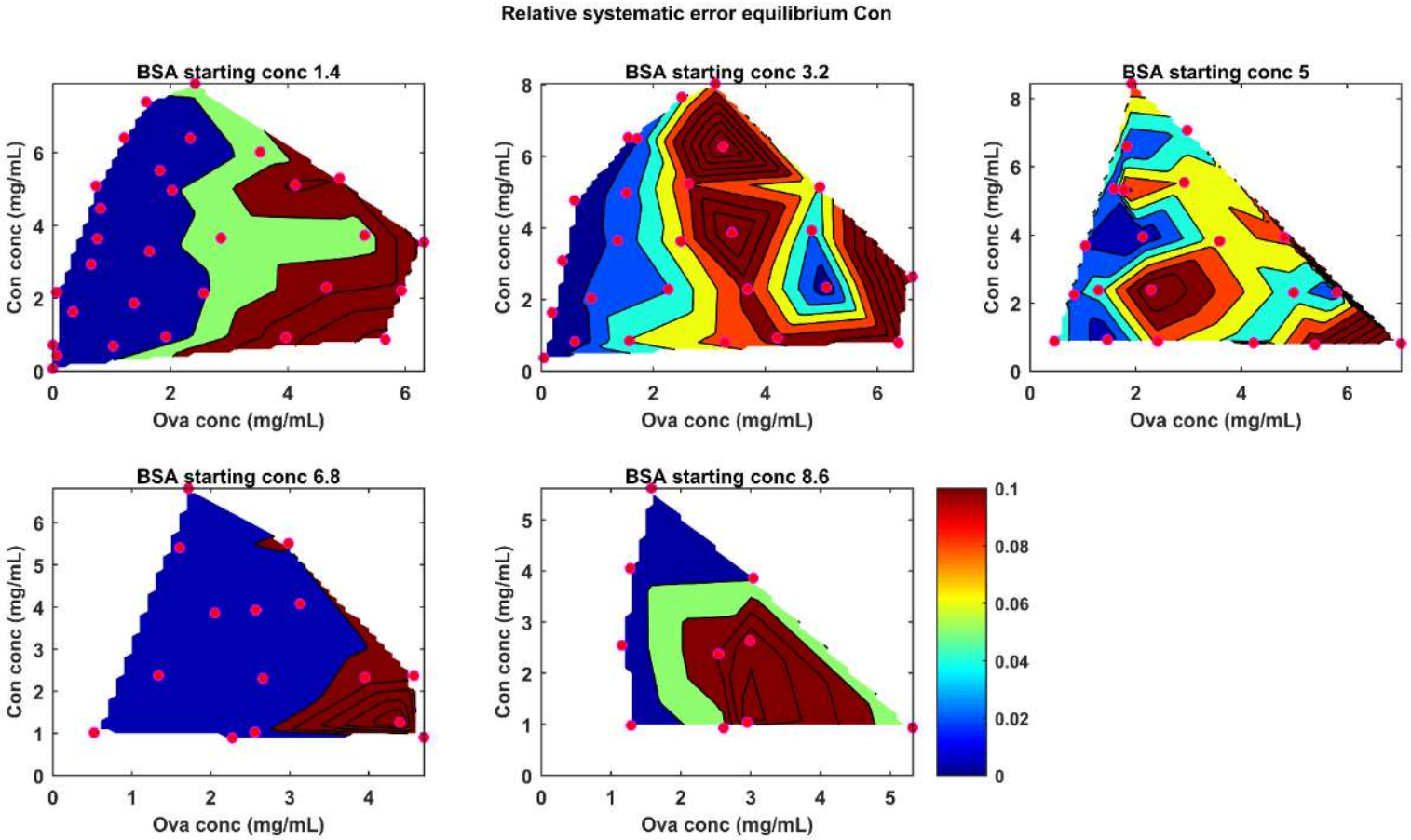

Figure 11-35 Contour plots showing propagated relative systematic error in q for Con in ternary isotherms at pH 9 25 mM NaCl. The colours represent relative adsorbed error. Each contour sub-plot can be thought of as an Ova-Con binary isotherm at different starting concentrations of BSA, the starting concentration of BSA in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of Ova and Con measured directly after outlier rejection.

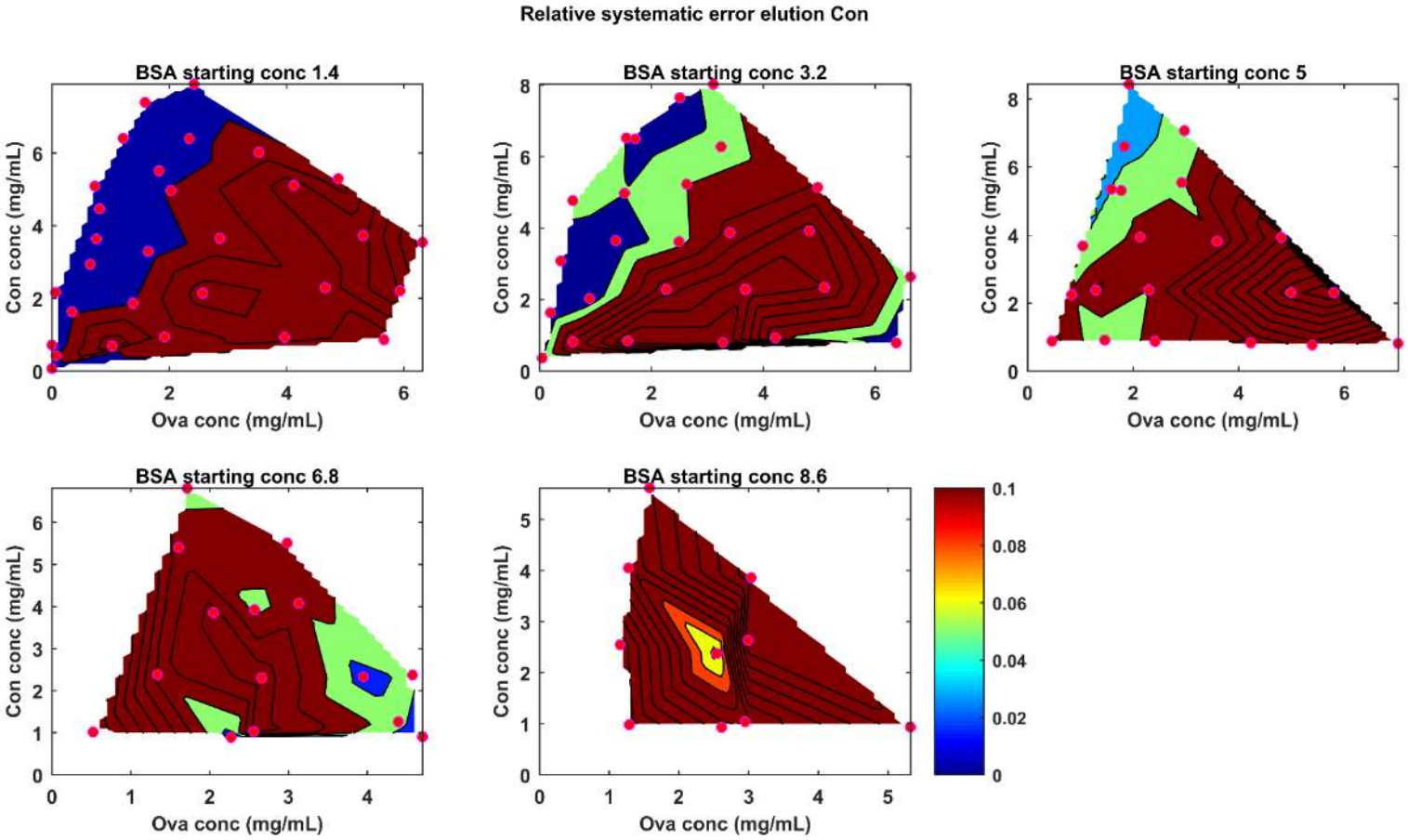

Figure 11-36 Contour plots showing propagated relative systematic error in q for Con in ternary isotherms at pH 9 25 mM NaCl. The colours represent relative adsorbed error. Each contour sub-plot can be thought of as an Ova-Con binary isotherm at different starting concentrations of BSA, the starting concentration of BSA in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of Ova and Con measured directly after outlier rejection.*

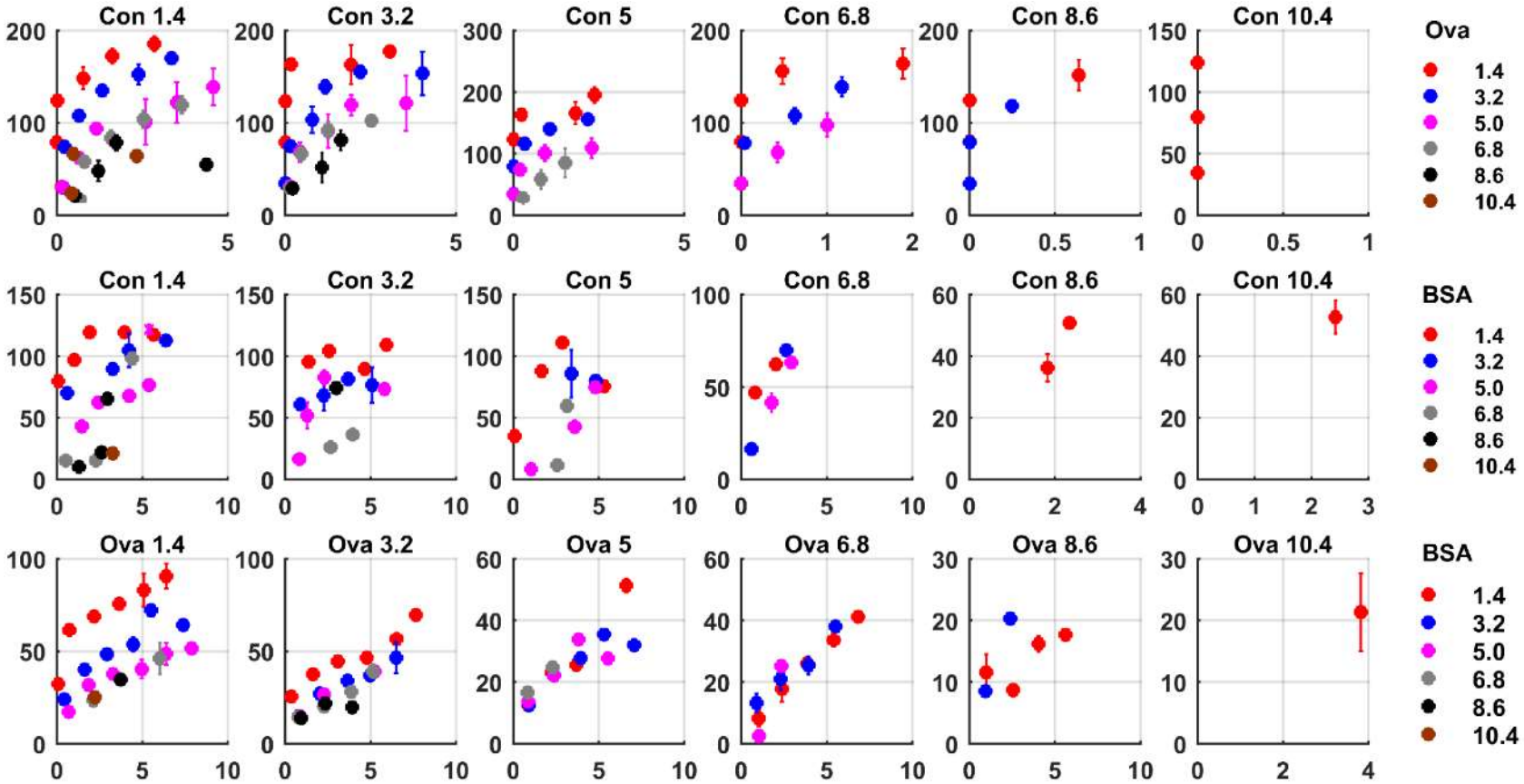

Figure 11-37 Ternary isotherm of BSA, Ova and Con plot as a series of single component isotherms at different starting concentrations of competitor protein at pH 9 and 25 mM NaCl. The plot titles display the starting concentration of one of the competitors and the legends display the starting concentration of the second competing protein in mg/mL. All x-axis display liquid concentrations in mg/mL and all y-axis represent adsorbed concentration in mg/mL of resin. Top row of 6 plots display BSA isotherms, second row of 6 plots displays Ova isotherms and the third row of 6 plots displays Con isotherms. Circles represent q data and x represents q. Error bars represent standard deviation of experimental data. The BSA isotherm only displays q data as q has lower levels of relative systametic error, the Ova and Con isotherms displays q and q* data after trimming as both q and q* datasets showed significant systematic relative error.*

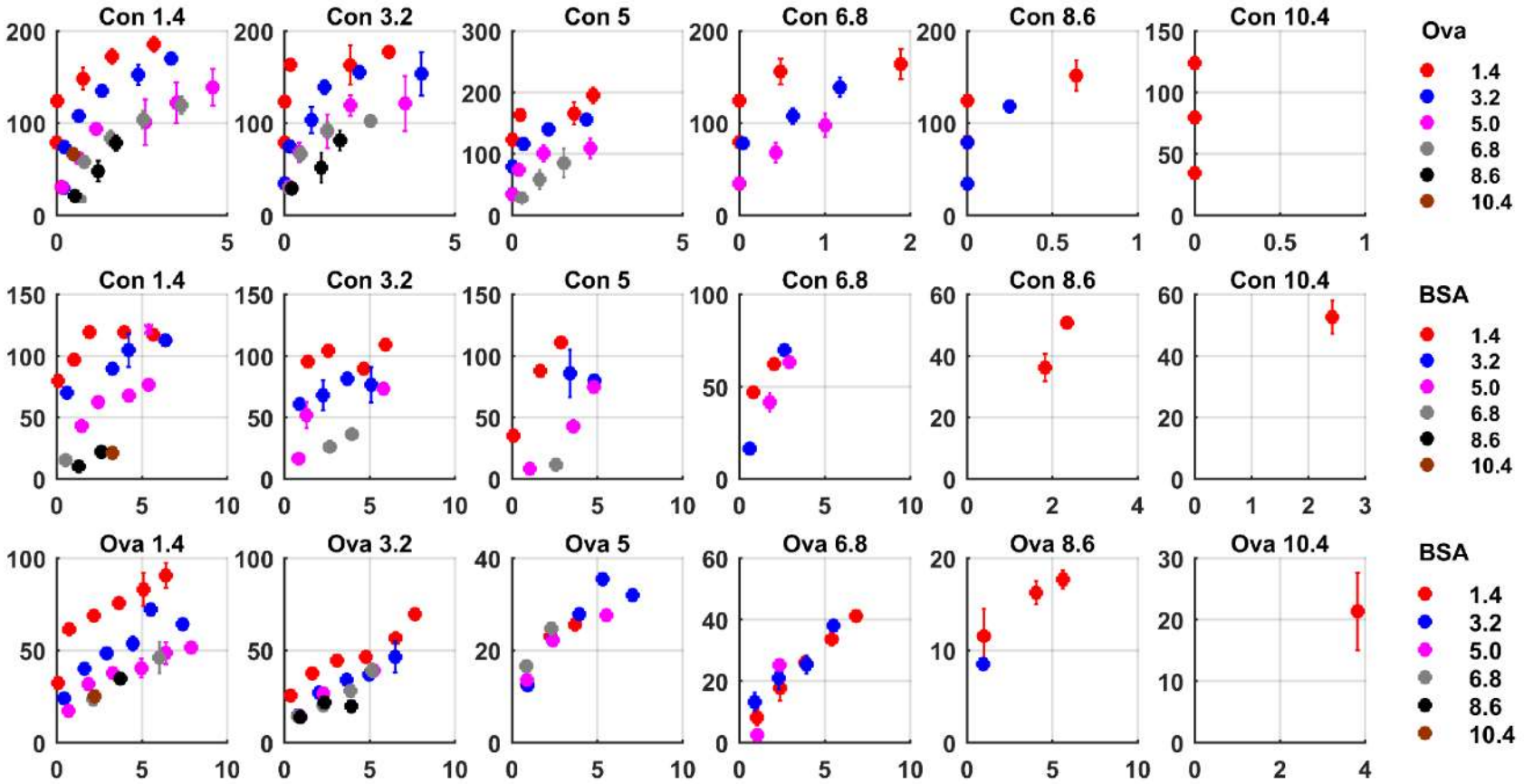

Figure 11-38 Ternary isotherm of BSA, Ova and Con plot as a series of single component isotherms at different starting concentrations of competitor protein at pH 9 and 25 mM NaCl. The plot titles display the starting concentration of one of the competitors and the legends display the starting concentration of the second competing protein in mg/mL. All x-axis display liquid concentrations in mg/mL and all y-axis represent adsorbed concentration in mg/mL of resin. Top row of 6 plots display BSA isotherms, second row of 6 plots displays Ova isotherms and the third row of 6 plots displays Con isotherms. Circles represent q data and x represents q. Error bars represent standard deviation of experimental data. The BSA isotherm only displays q data after manually omitting points, the Ova isotherm displays q and q* data after trimming and manually omitting points, the Con isotherm displays q and q* data after trimming and manually omitting points. As implausible isotherm behaviour was observed after previous steps so manual omission was applied to the isotherm data.*

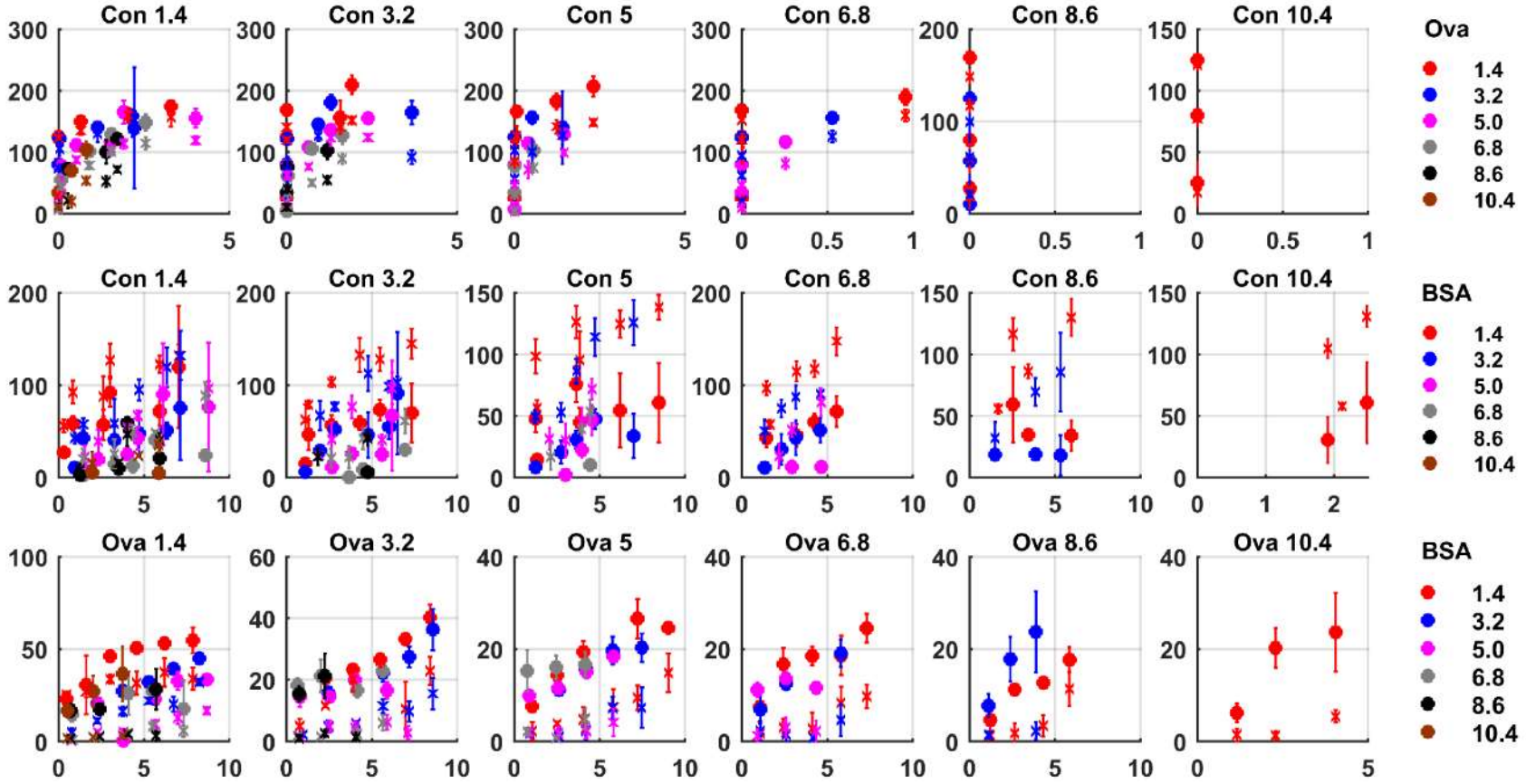

Figure 11-39 Ternary isotherm of BSA, Ova and Con plot as a series of single component isotherms at different starting concentrations of competitor protein at pH 9 and 50 mM NaCl. The plot titles display the starting concentration of one of the competitors and the legends display the starting concentration of the second competing protein in mg/mL. All x-axis display liquid concentrations in mg/mL and all y-axis represent adsorbed concentration in mg/mL of resin. Top row of 6 plots display BSA isotherms, second row of 6 plots displays Ova isotherms and the third row of 6 plots displays Con isotherms. Circles represent q data and x represents q. Error bars represent standard deviation of experimental data. As poor agreement is observed between q and q* measures propagation of systematic errors for BSA, Ova and Con are displayed. Additionally, large error bars are observed for BSA and Ova isotherms and so propagation of both systematic random errors is included from Figure 11-40 to Figure 11-52.*

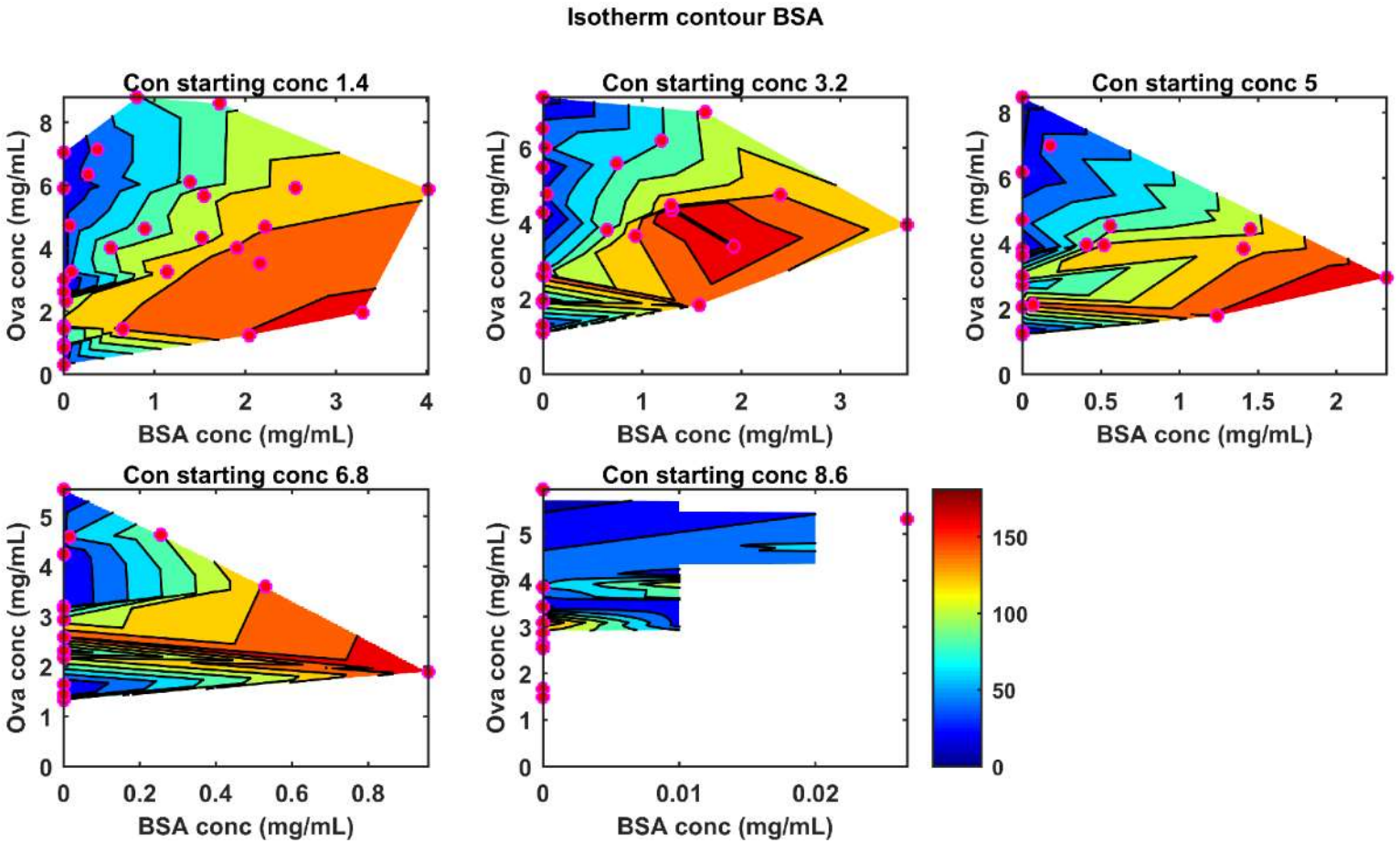

Figure 11-40 Ternary isotherm at pH 9 50 mM NaCl, contours display adsorbed concentration of BSA, adsorbed concentrations is as an average q and q. Isotherm has been drawn as a contour plot as a reference for error propagation plots. Each contour sub-plot can be thought of as a BSA-Ova binary isotherm at different starting concentrations of Con, the starting concentration of Con in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of BSA and Ova measured directly after outlier rejection.*

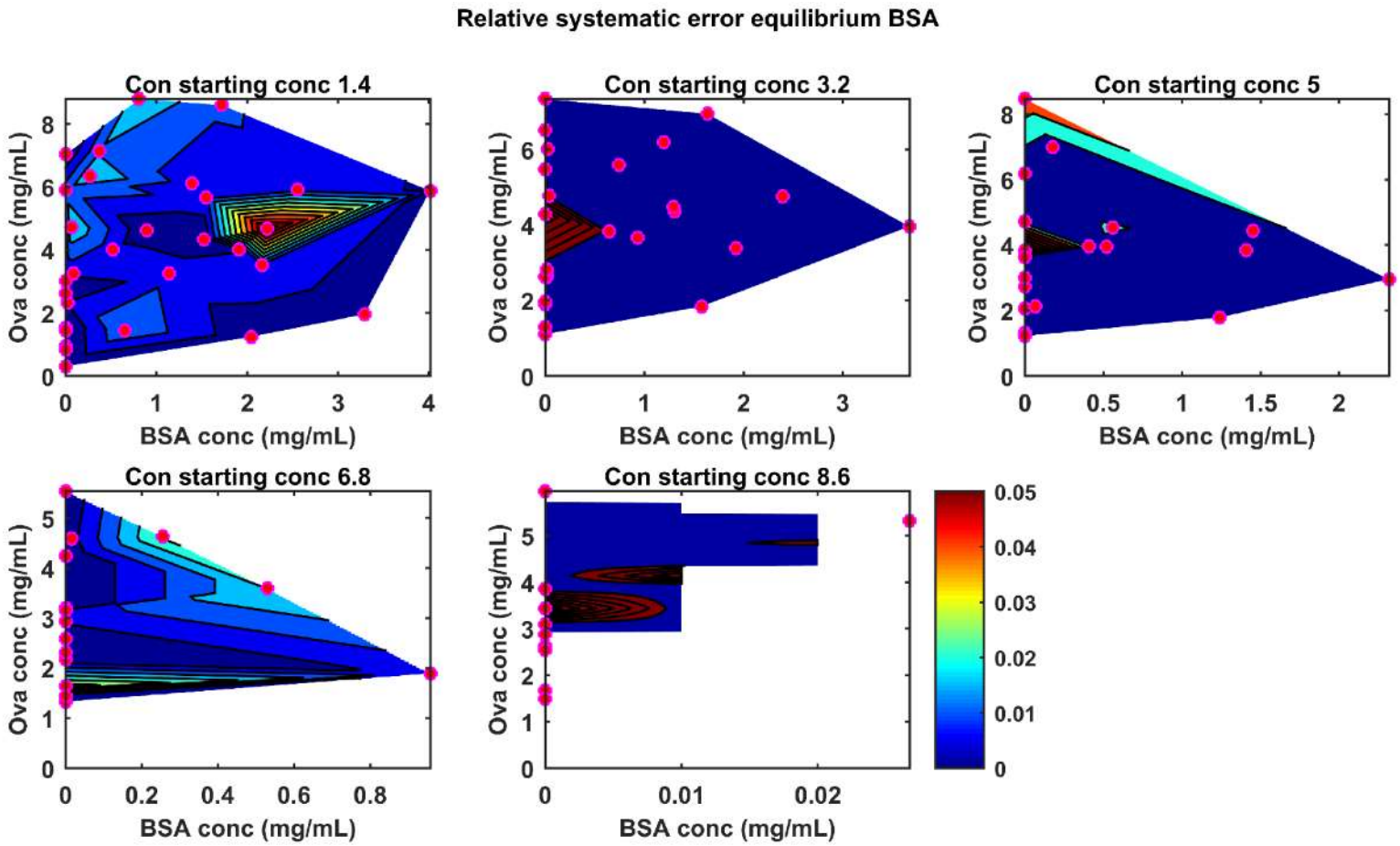

Figure 11-41 Contour plots showing propagated relative systematic error in q for BSA in ternary isotherms at pH 9 50 mM NaCl. The colours represent relative adsorbed error. Each contour sub-plot can be thought of as a BSA-Ova binary isotherm at different starting concentrations of Con, the starting concentration of Con in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of BSA and Ova measured directly after outlier rejection.

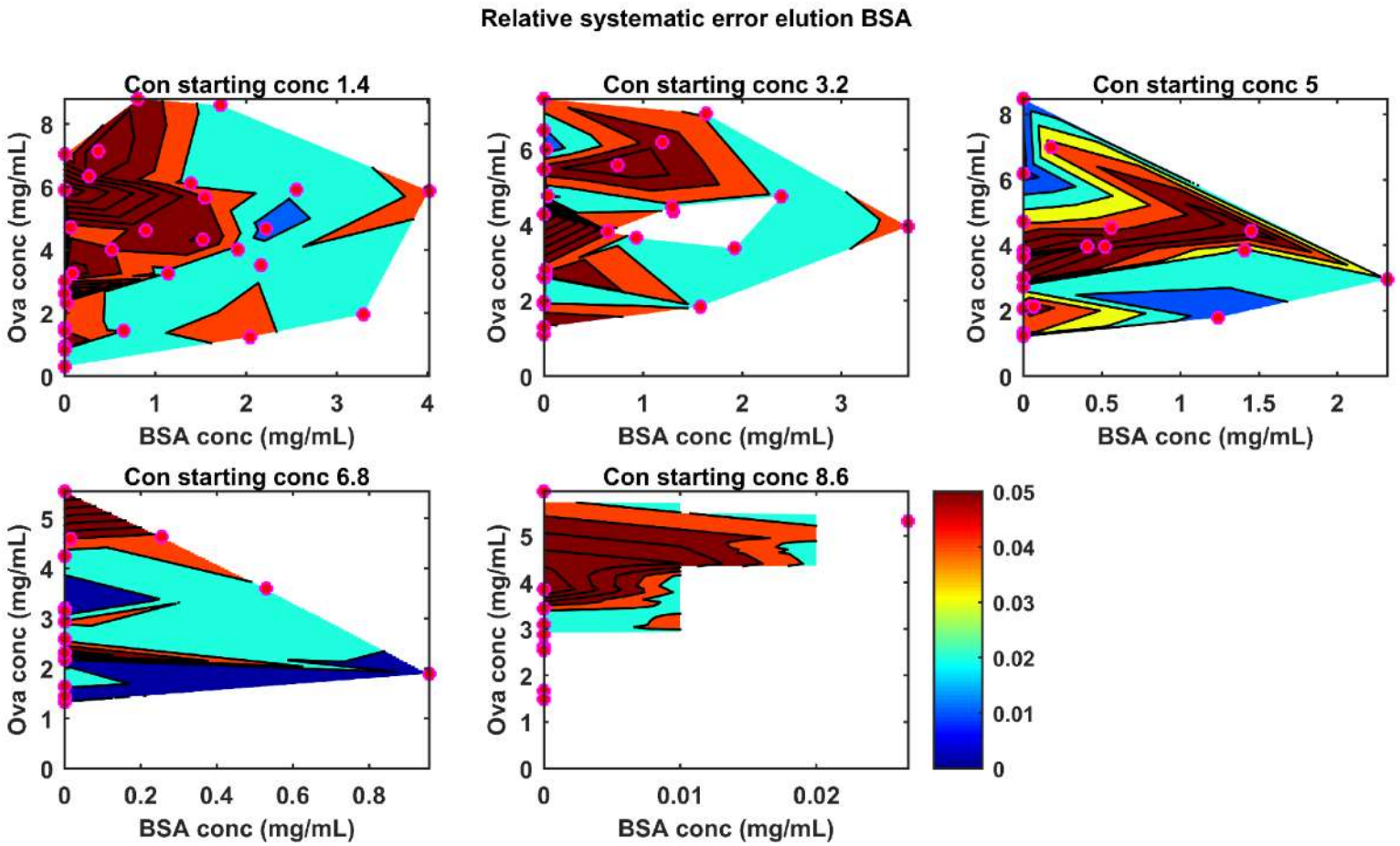

Figure 11-42 Contour plots showing propagated relative systematic error in q for BSA in ternary isotherms at pH 9 50 mM NaCl. The colours represent relative adsorbed error. Each contour sub-plot can be thought of as a BSA-Ova binary isotherm at different starting concentrations of Con, the starting concentration of Con in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of BSA and Ova measured directly after outlier rejection.*

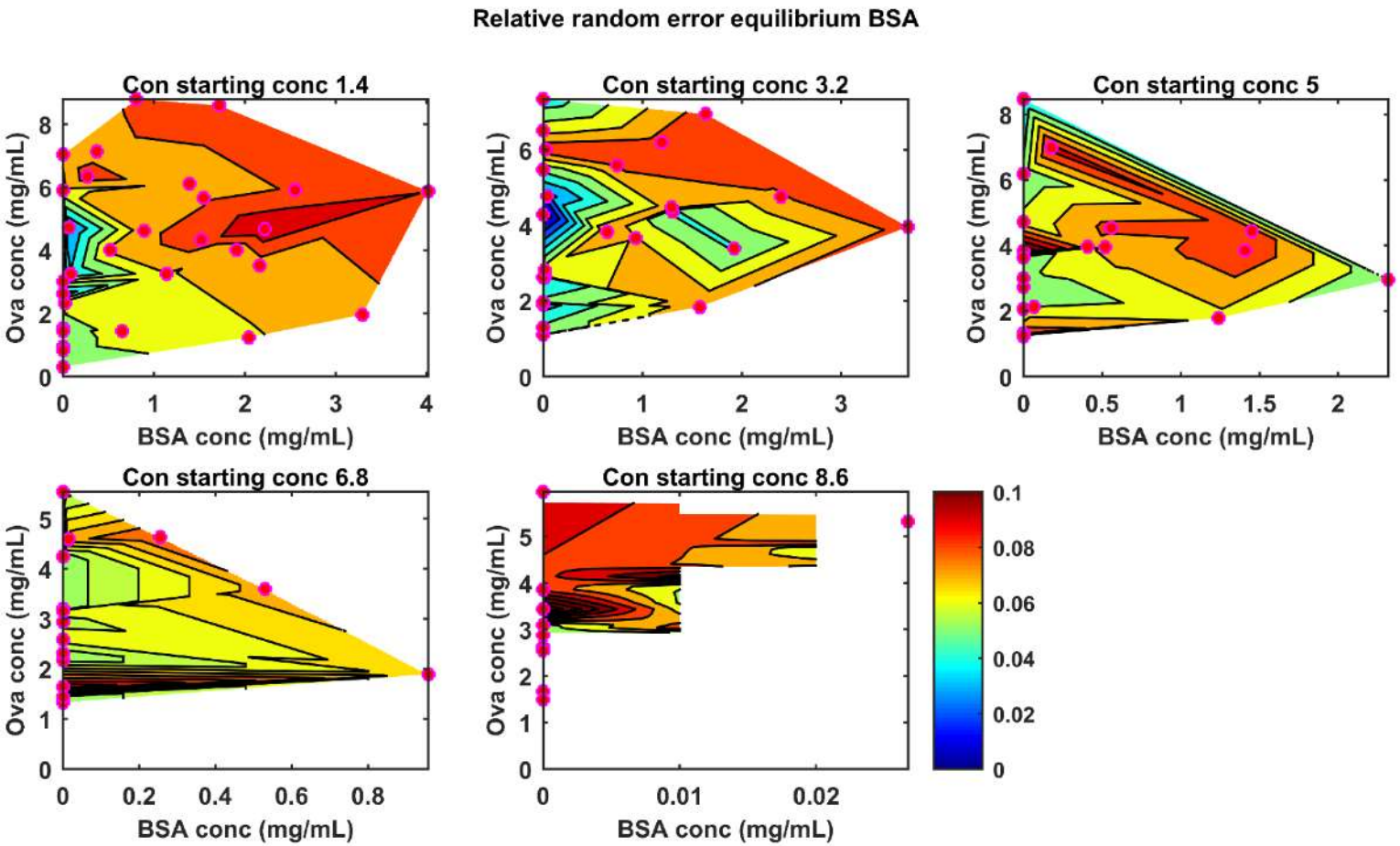

Figure 11-43 Contour plots showing propagated relative random error in q for BSA in ternary isotherms at pH 9 50 mM NaCl. The colours represent relative adsorbed error. Each contour sub-plot can be thought of as a BSA-Ova binary isotherm at different starting concentrations of Con, the starting concentration of Con in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of BSA and Ova measured directly after outlier rejection.

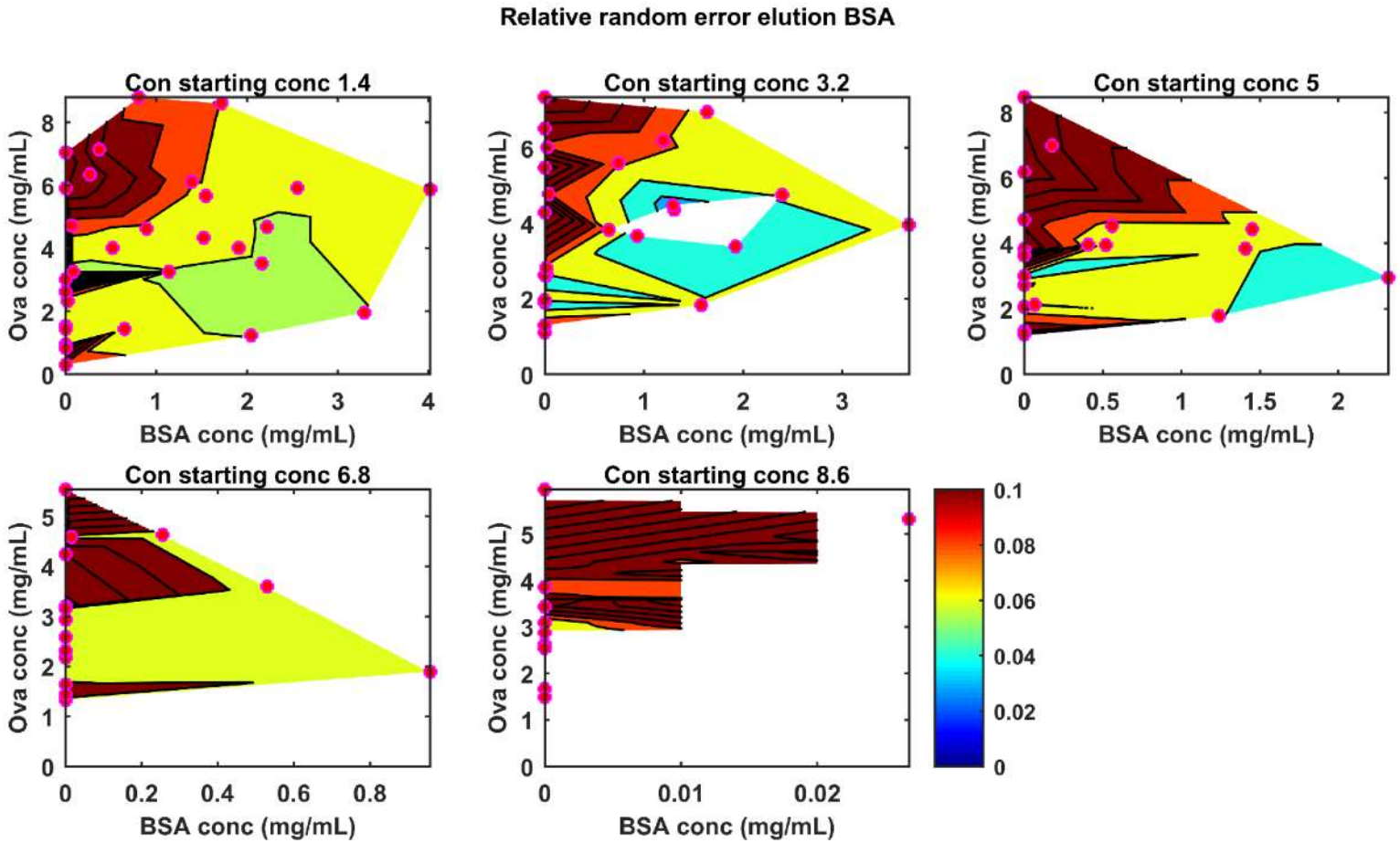

Figure 11-44 Contour plots showing propagated relative random error in q for BSA in ternary isotherms at pH 9 50 mM NaCl. The colours represent relative adsorbed error. Each contour sub-plot can be thought of as a BSA-Ova binary isotherm at different starting concentrations of Con, the starting concentration of Con in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of BSA and Ova measured directly after outlier rejection.*

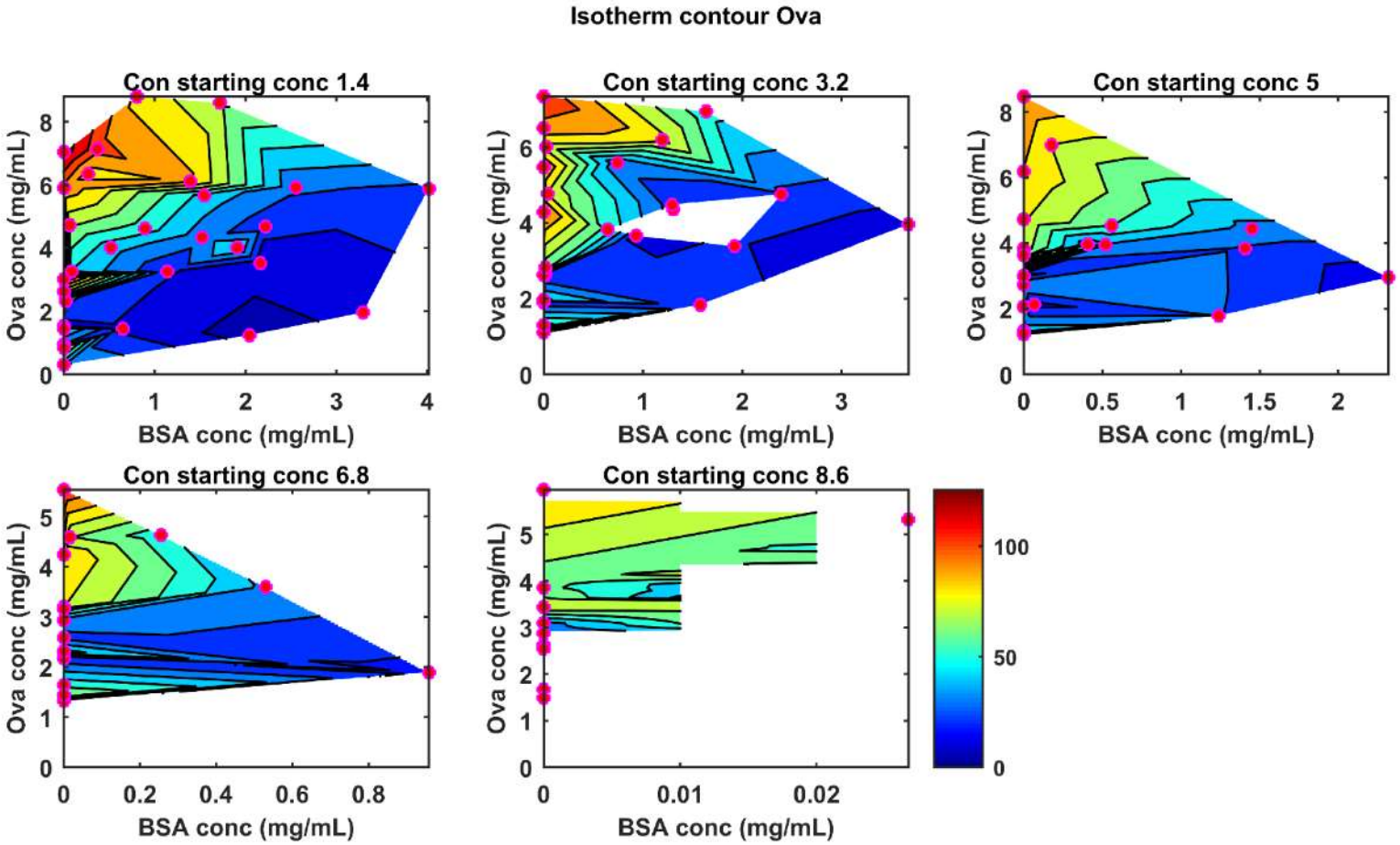

Figure 11-45 Ternary isotherm at pH 9 50 mM NaCl, contours display adsorbed concentration of Ova, adsorbed concentrations is as an average q and q. Isotherm has been drawn as a contour plot as a reference for error propagation plots. Each contour sub-plot can be thought of as a BSA-Ova binary isotherm at different starting concentrations of Con, the starting concentration of Con in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of BSA and Ova measured directly after outlier rejection.*

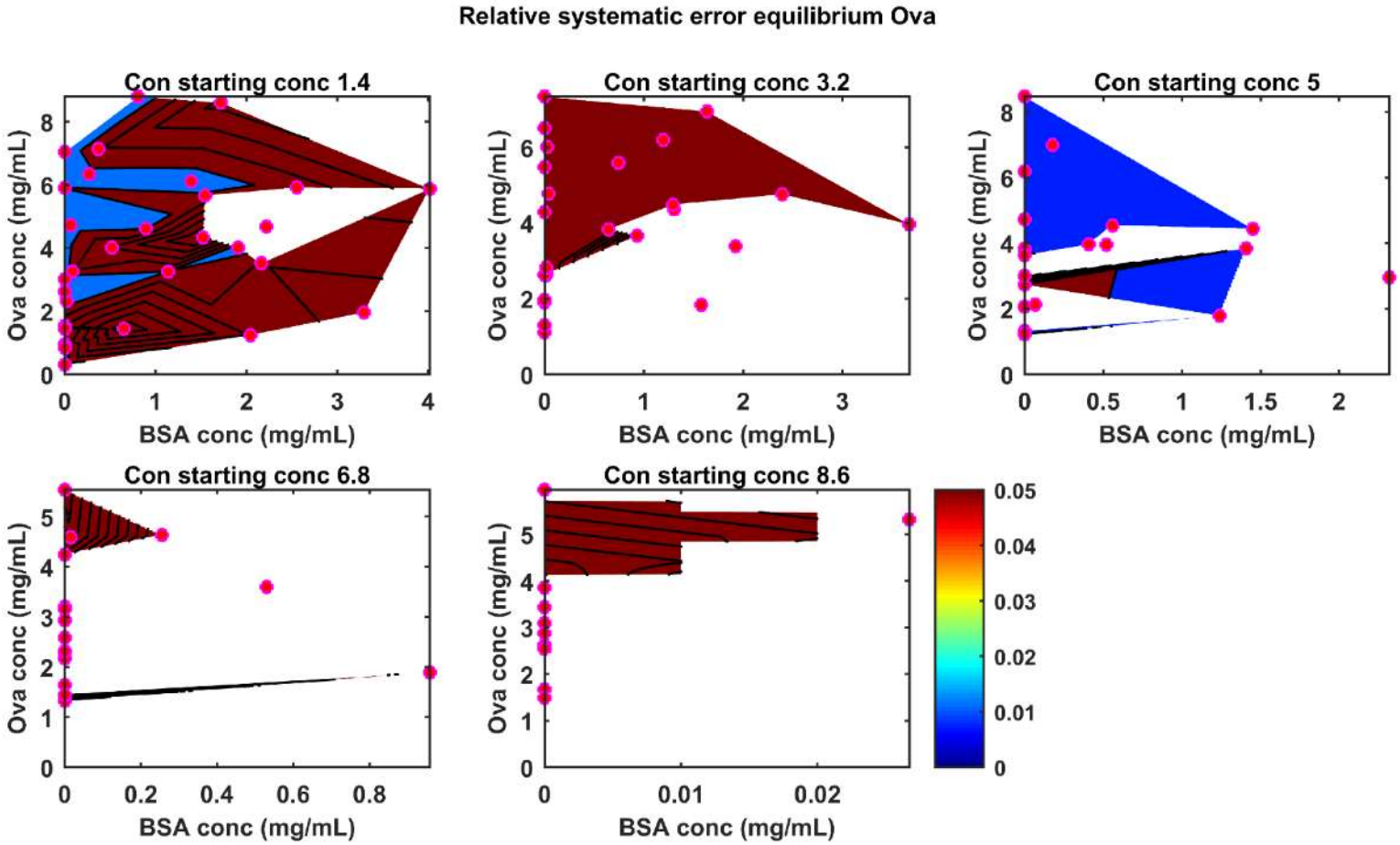

Figure 11-46 Contour plots showing propagated relative systematic error in q for Ova in ternary isotherms at pH 9 50 mM NaCl. The colours represent relative adsorbed error. Each contour sub-plot can be thought of as a BSA-Ova binary isotherm at different starting concentrations of Con, the starting concentration of Con in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of BSA and Ova measured directly after outlier rejection.

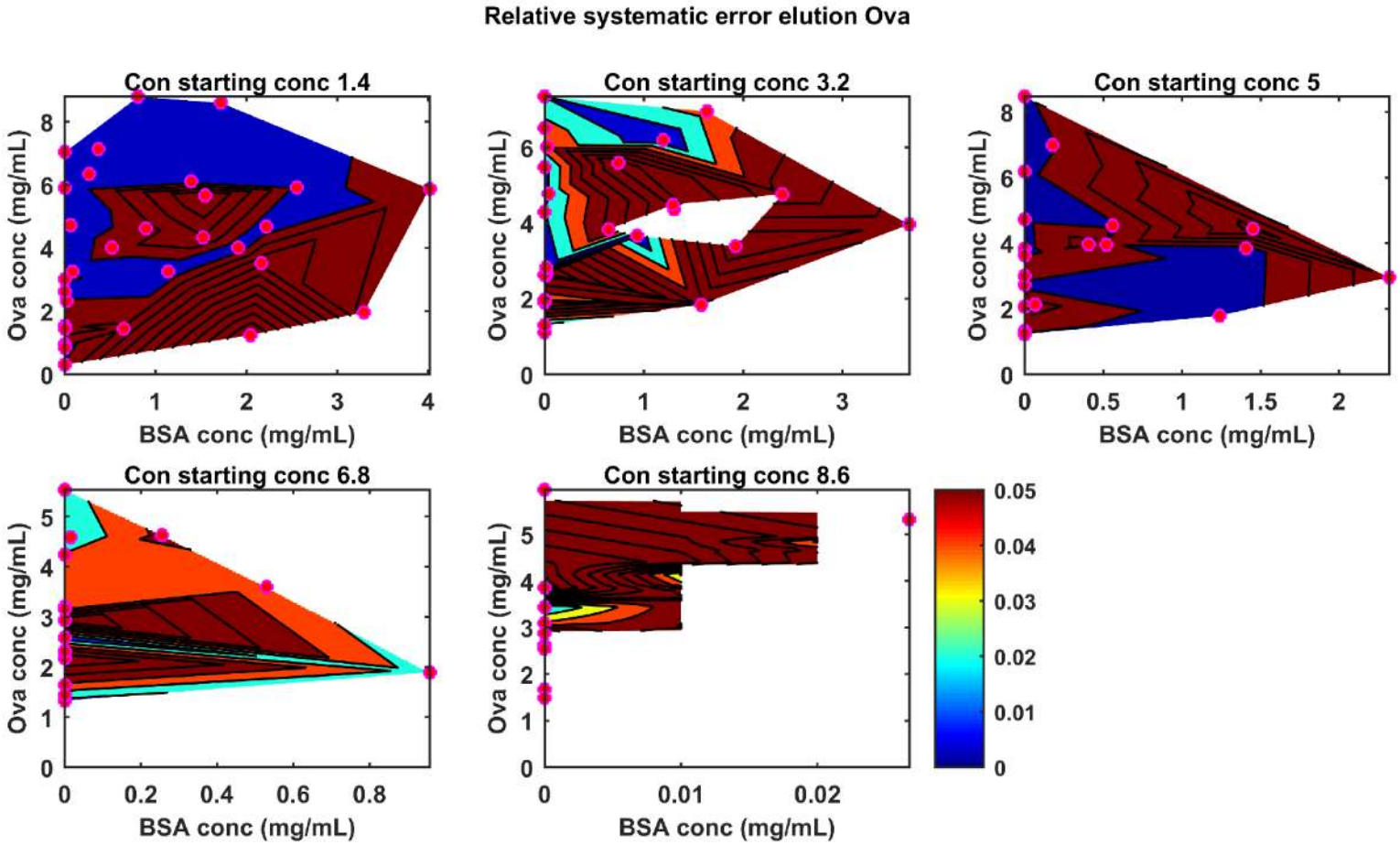

Figure 11-47 Contour plots showing propagated relative systematic error in q for Ova in ternary isotherms at pH 9 50 mM NaCl. The colours represent relative adsorbed error. Each contour sub-plot can be thought of as a BSA-Ova binary isotherm at different starting concentrations of Con, the starting concentration of Con in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of BSA and Ova measured directly after outlier rejection.*

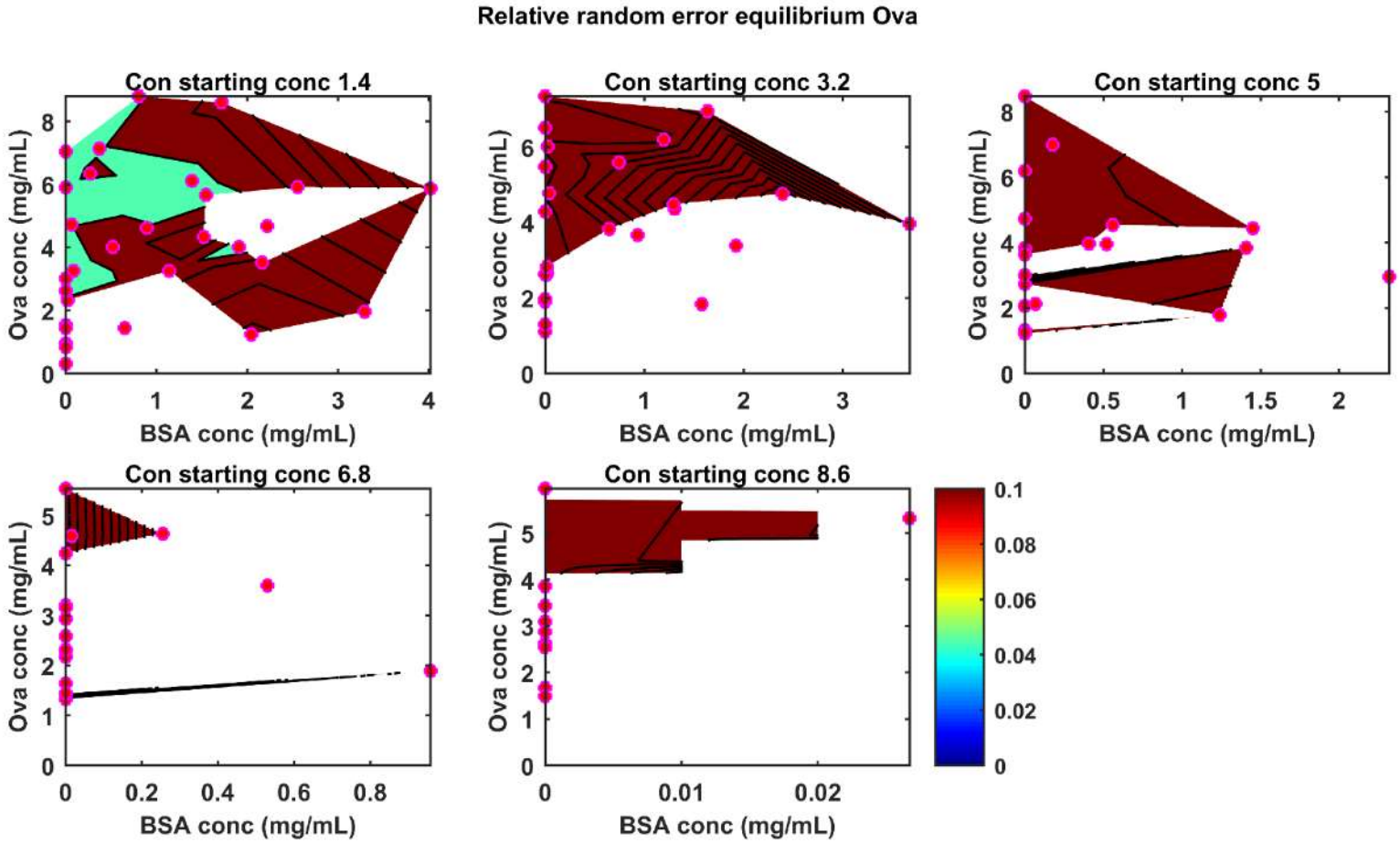

Figure 11-48 Contour plots showing propagated relative random error in q for Ova in ternary isotherms at pH 9 50 mM NaCl. The colours represent relative adsorbed error. Each contour sub-plot can be thought of as a BSA-Ova binary isotherm at different starting concentrations of Con, the starting concentration of Con in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of BSA and Ova measured directly after outlier rejection.

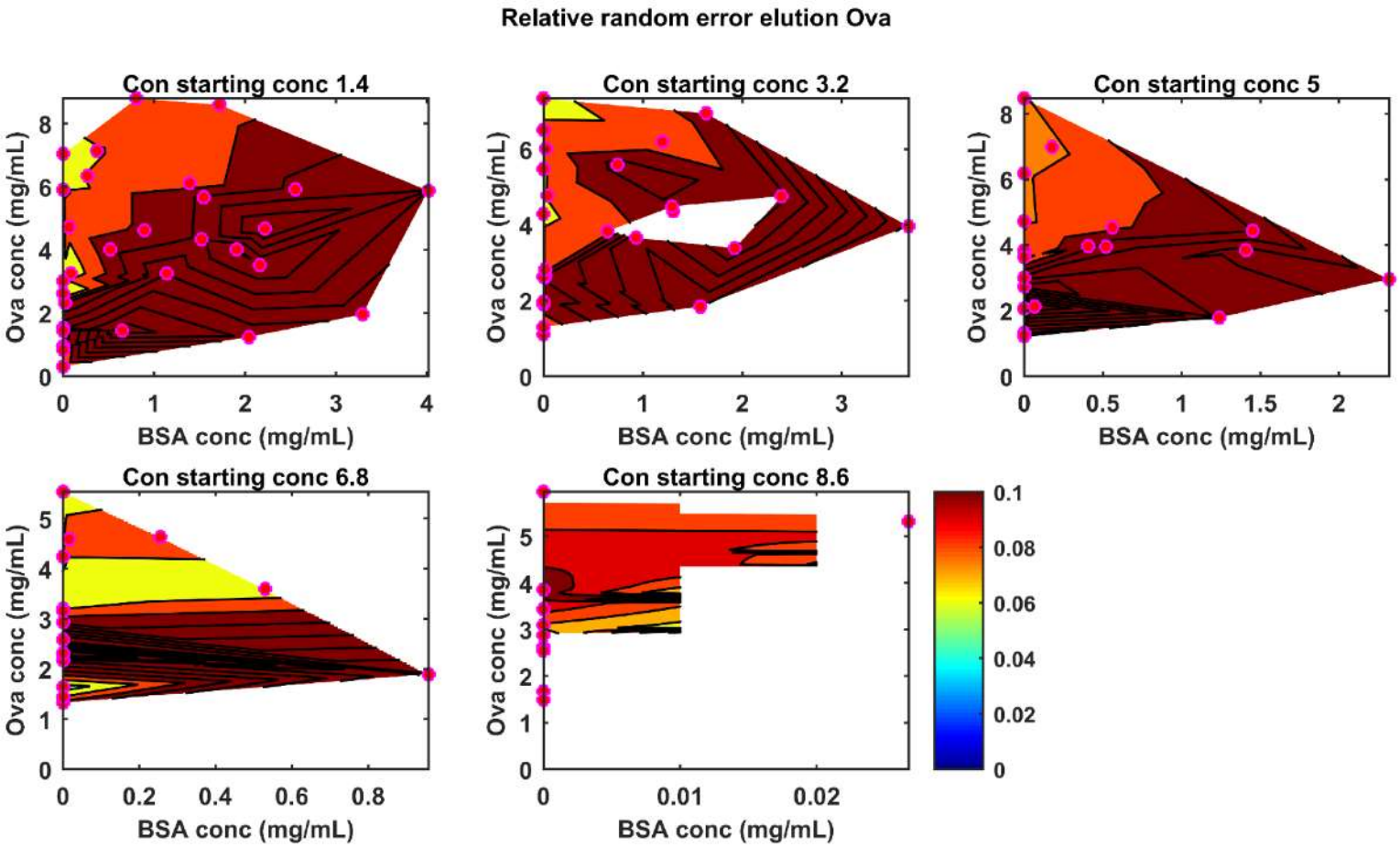

Figure 11-49 Contour plots showing propagated relative random error in q for Ova in ternary isotherms at pH 9 50 mM NaCl. The colours represent relative adsorbed error. Each contour sub-plot can be thought of as a BSA-Ova binary isotherm at different starting concentrations of Con, the starting concentration of Con in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of BSA and Ova measured directly after outlier rejection.*

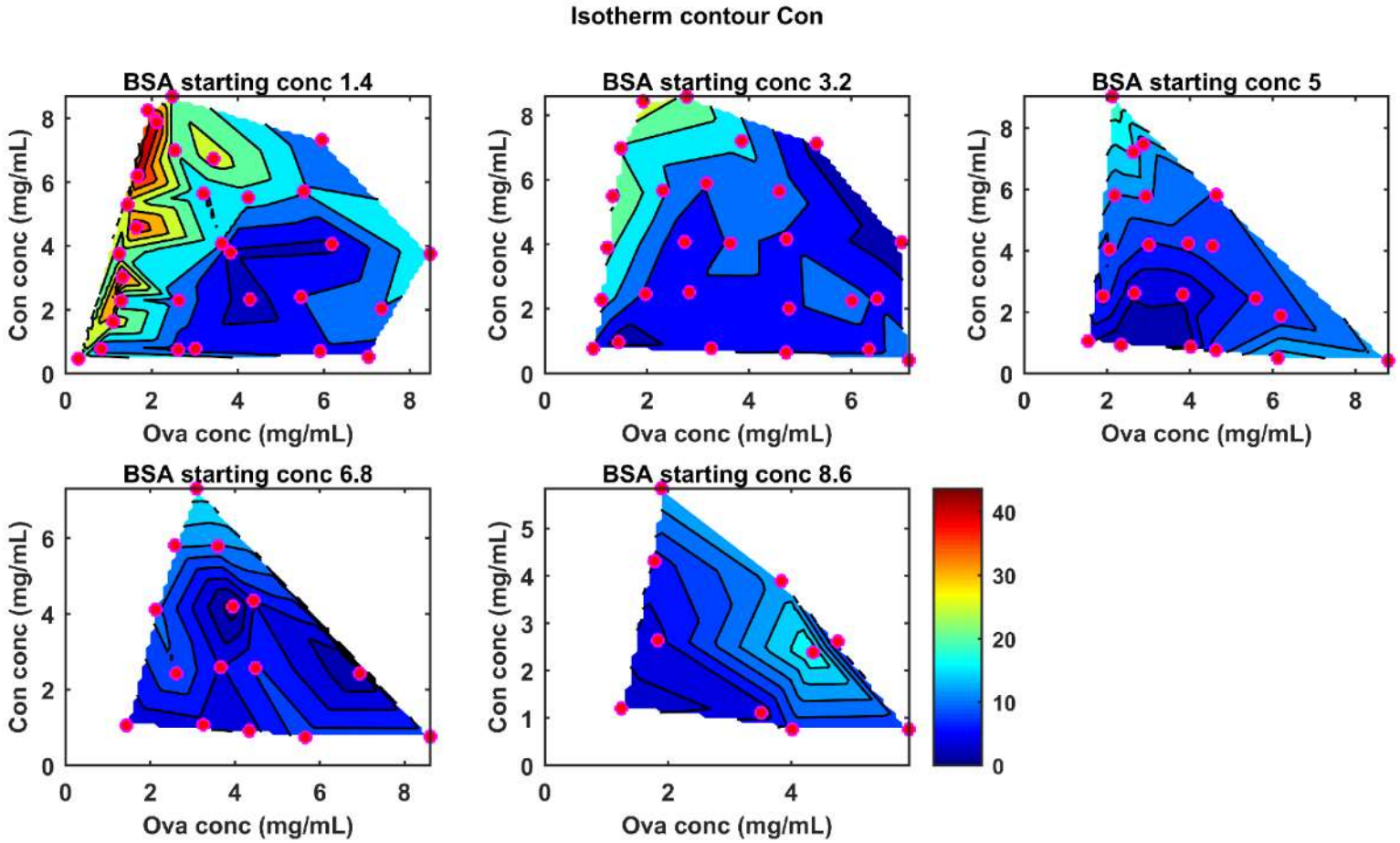

Figure 11-50 Ternary isotherm at pH 9 50 mM NaCl, contours display adsorbed concentration of Con, adsorbed concentrations is as an average q and q. Isotherm has been drawn as a contour plot as a reference for error propagation plots. Each contour sub-plot can be thought of as an Ova-Con binary isotherm at different starting concentrations of BSA, the starting concentration of BSA in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of Ova and Con measured directly after outlier rejection.*

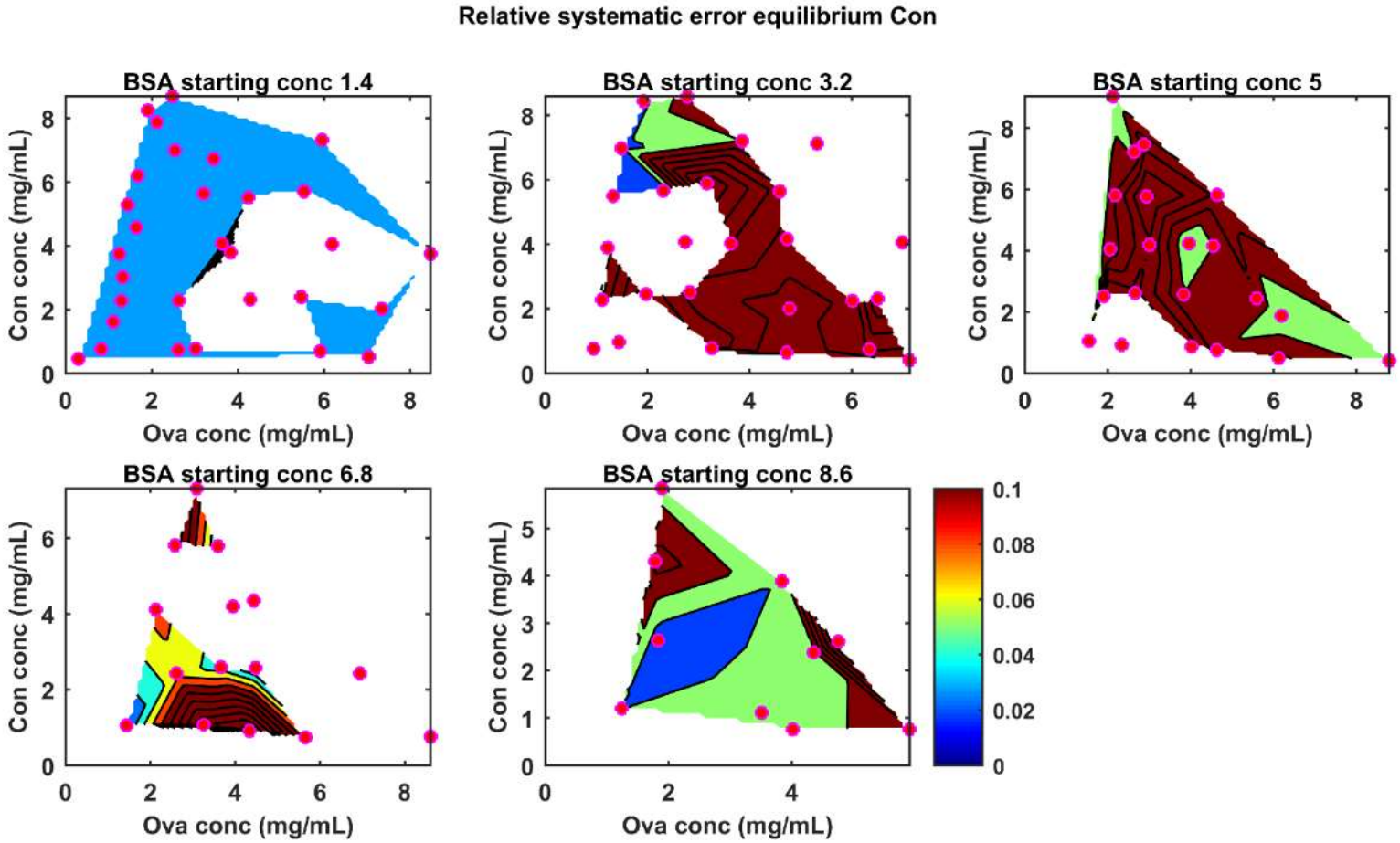

Figure 11-51 Contour plots showing propagated relative systematic error in q for Con in ternary isotherms at pH 9 50 mM NaCl. The colours represent relative adsorbed error. Each contour sub-plot can be thought of as an Ova-Con binary isotherm at different starting concentrations of BSA, the starting concentration of BSA in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of Ova and Con measured directly after outlier rejection.

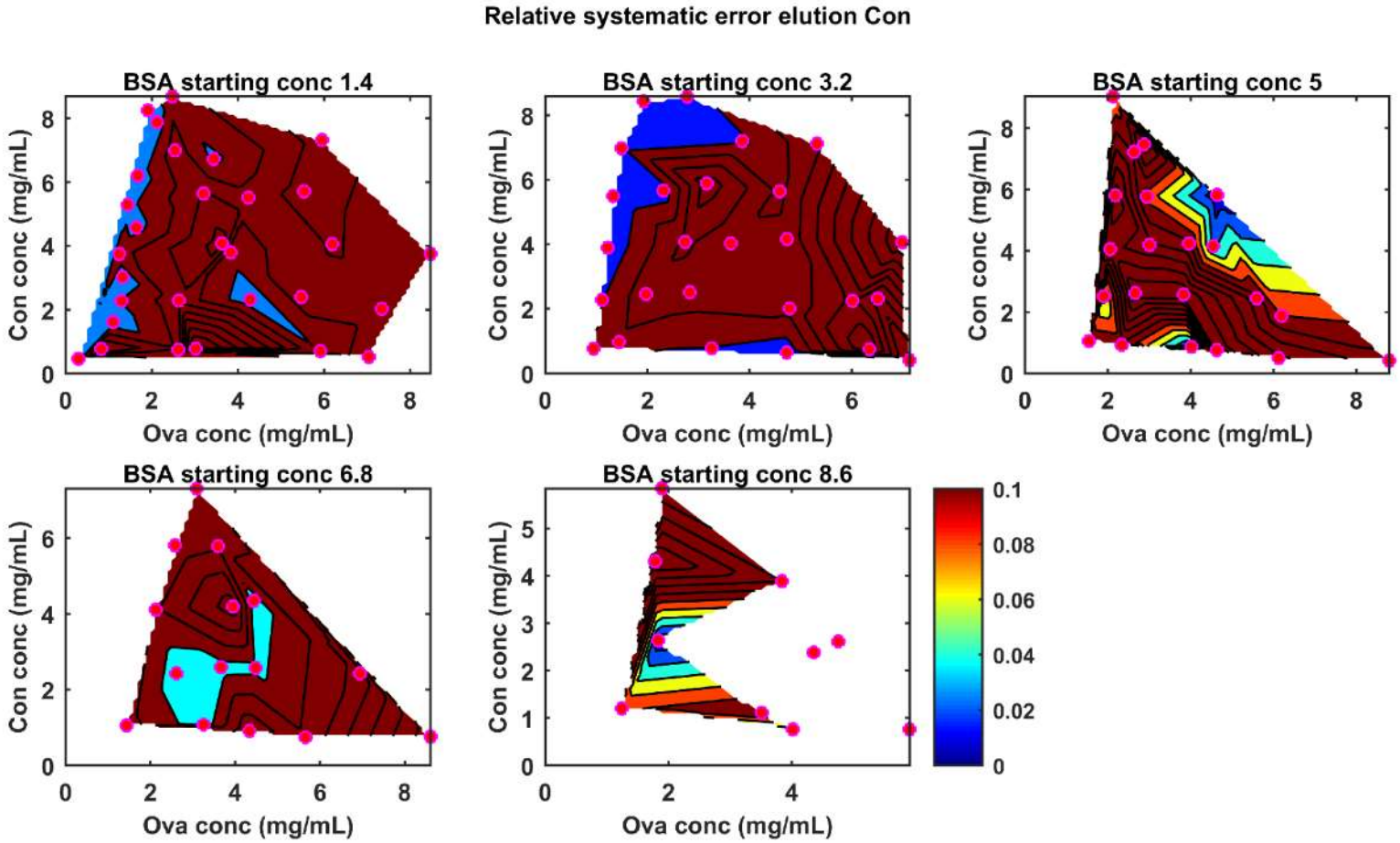

Figure 11-52 Contour plots showing propagated relative systematic error in q for Con in ternary isotherms at pH 9 50 mM NaCl. The colours represent relative adsorbed error. Each contour sub-plot can be thought of as an Ova-Con binary isotherm at different starting concentrations of BSA, the starting concentration of BSA in mg/mL is the title of each sub-plot. x and y-axis display average equilibrium concentrations of Ova and Con measured directly after outlier rejection.*

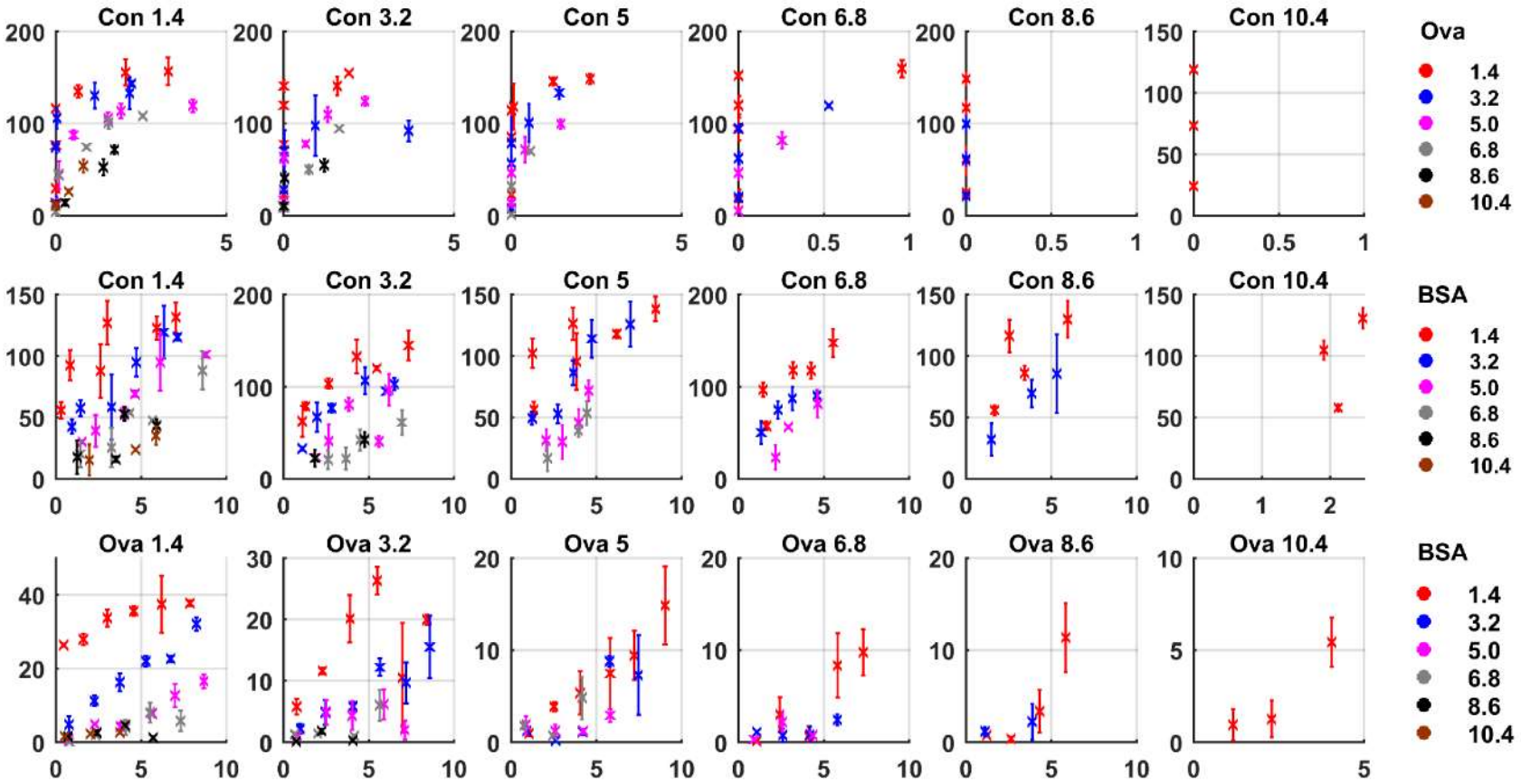

Figure 11-53 Ternary isotherm of BSA, Ova and Con plot as a series of single component isotherms at different starting concentrations of competitor protein at pH 9 and 50 mM NaCl. The plot titles display the starting concentration of one of the competitors and the legends display the starting concentration of the second competing protein in mg/mL. All x-axis display liquid concentrations in mg/mL and all y-axis represent adsorbed concentration in mg/mL of resin. Top row of 6 plots display BSA isotherms, second row of 6 plots displays Ova isotherms and the third row of 6 plots displays Con isotherms. Only q data has been plotted. Error bars represent standard deviation of experimental data. Only q* data is observed because q data either expressed negative adsorption values or implausible trends.*

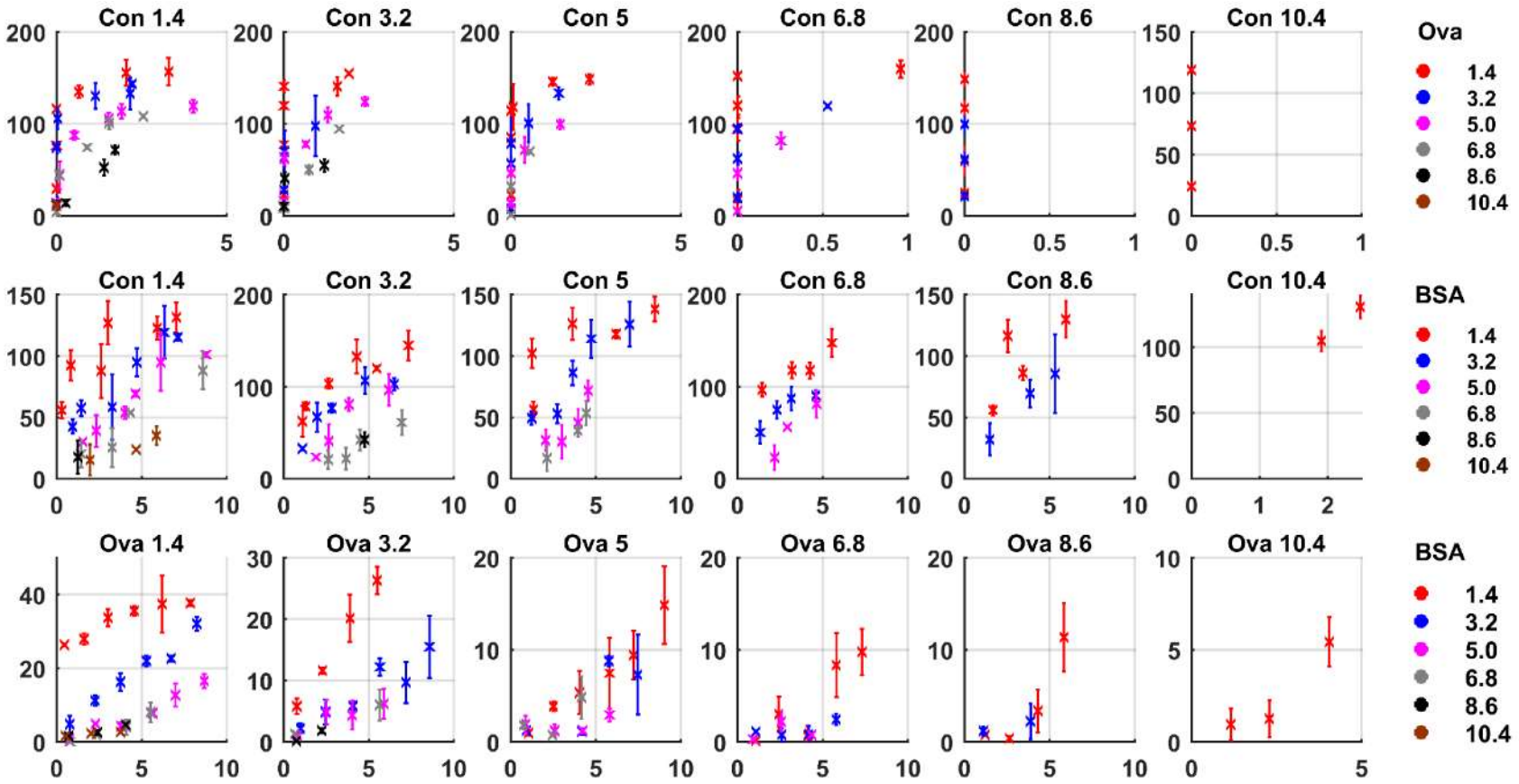

*Figure 11-54 Ternary isotherm of BSA, Ova and Con plot as a series of single component isotherms at different starting concentrations of competitor protein at pH 9 and 50 mM NaCl. The plot titles display the starting concentration of one of the competitors and the legends display the starting concentration of the second competing protein in mg/mL. All x-axis display liquid concentrations in mg/mL and all y-axis represent adsorbed concentration in mg/mL of resin. Top row of 6 plots display BSA isotherms, second row of 6 plots displays Ova isotherms and the third row of 6 plots displays Con isotherms. Only q*data has been plotted after omitting some data points manually. Error bars represent standard deviation of experimental data. Points were manually omitted as the q* data set still displayed implausible behaviour after previous steps.*

11.3 Appendix C: Additional isotherm fitting data

Binary and ternary isotherms were fit to FLJ, SMA and Padé isotherm descriptions, fits not discussed in Chapter 8 Fitting multicomponent isotherm descriptions have been included here.

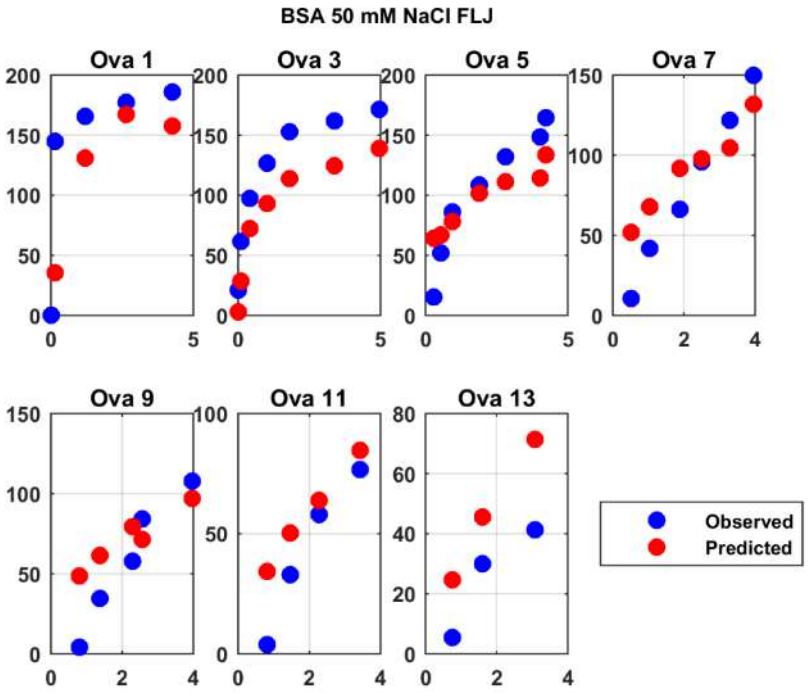

Figure 11-55 Freundlich-Langmuir-Jovanovic (FLJ) fit to binary BSA-Ova binary isotherm showing observed and predicted values for BSA adsorbed concentration at 50 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of Ova displayed as the title of each subplot.

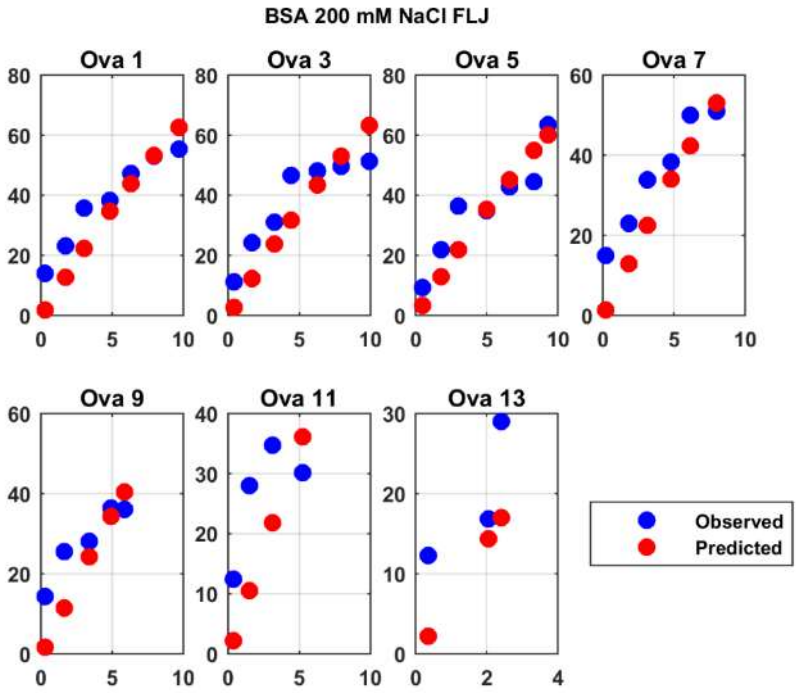

Figure 11-56 Freundlich-Langmuir-Jovanovic (FLJ) fit to binary BSA-Ova binary isotherm showing observed and predicted values for BSA adsorbed concentration at 200 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of Ova displayed as the title of each subplot.

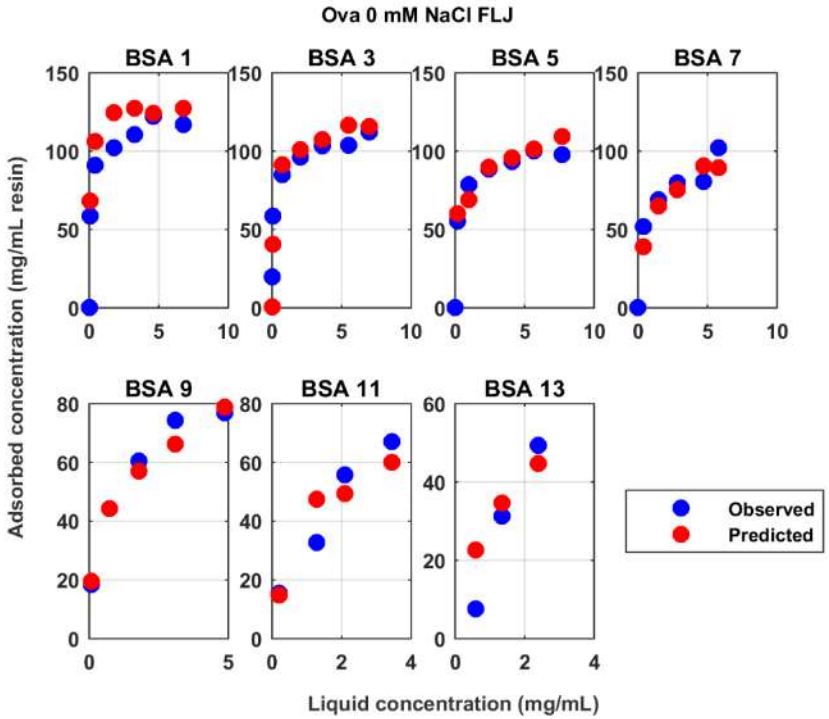

Figure 11-57 Freundlich-Langmuir-Jovanovic (FLJ) fit to binary BSA-Ova binary isotherm showing observed and predicted values for Ova adsorbed concentration at 0 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of BSA displayed as the title of each subplot.

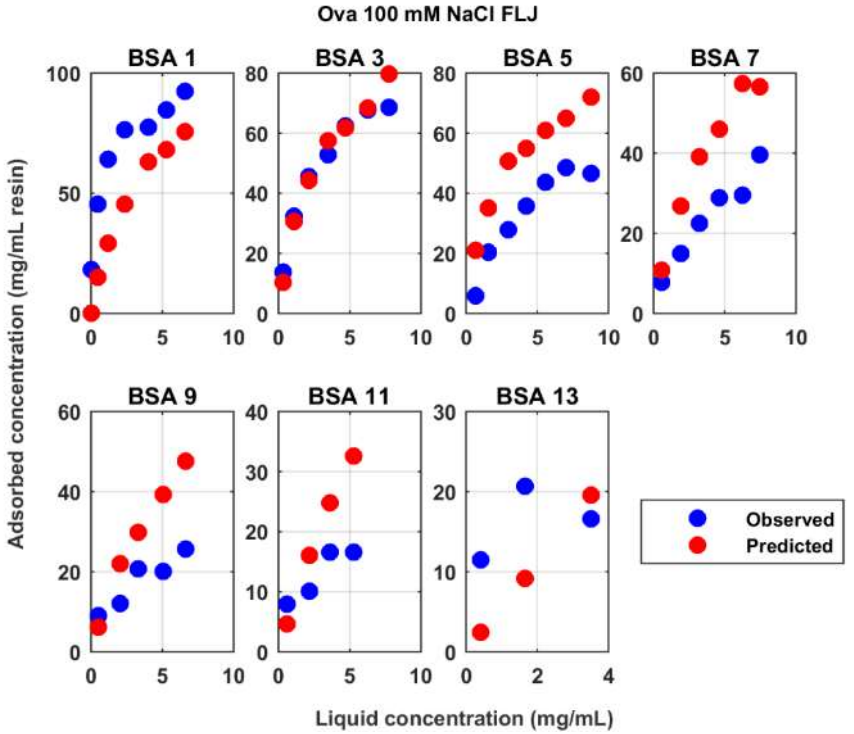

Figure 11-58 Freundlich-Langmuir-Jovanovic (FLJ) fit to binary BSA-Ova binary isotherm showing observed and predicted values for Ova adsorbed concentration at 100 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of BSA displayed as the title of each subplot.

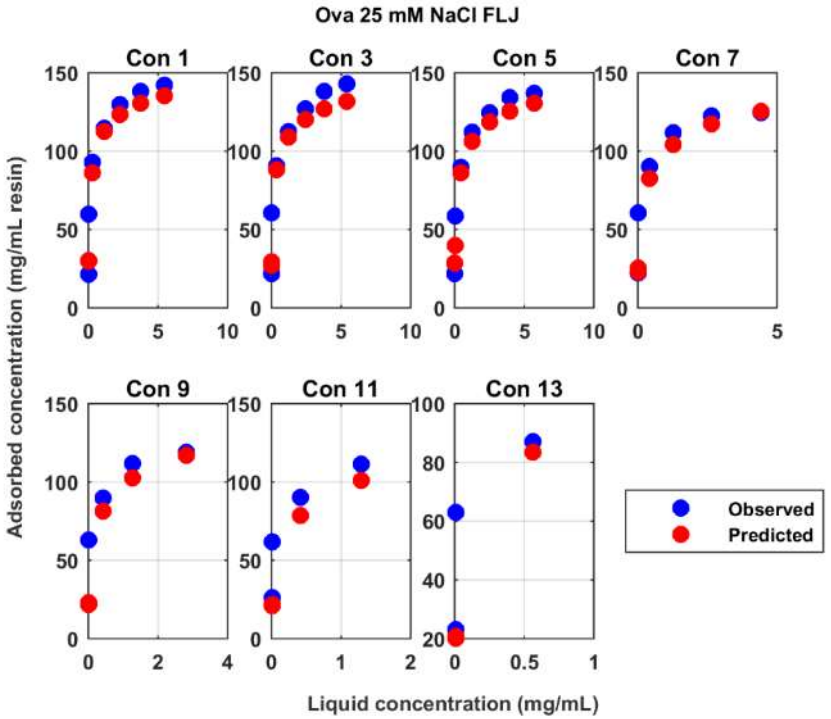

Figure 11-59 Freundlich-Langmuir-Jovanovic (FLJ) fit to binary Ova-Con binary isotherm showing observed and predicted values for Ova adsorbed concentration at 25 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of Con displayed as the title of each subplot.

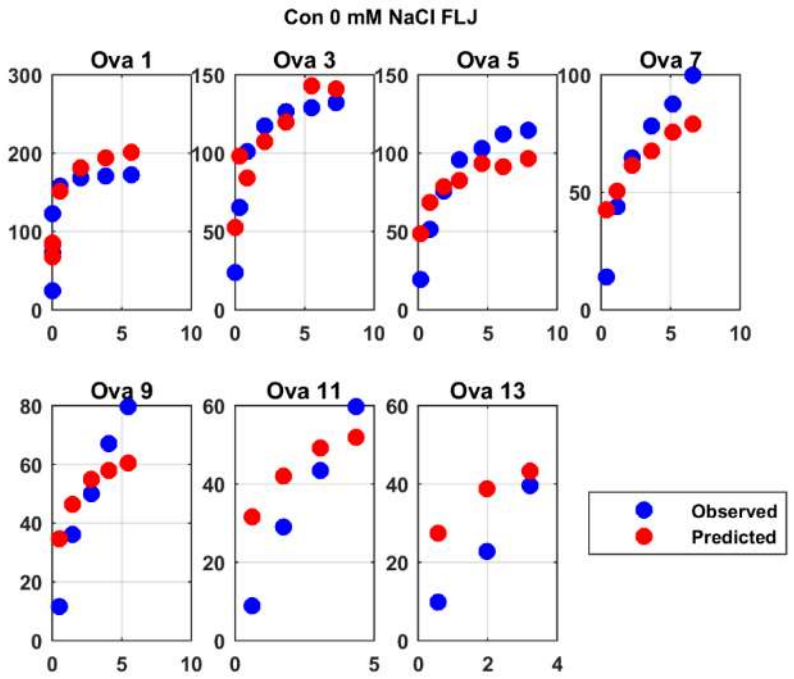

Figure 11-60 Freundlich-Langmuir-Jovanovic (FLJ) fit to binary Ova-Con binary isotherm showing observed and predicted values for Con adsorbed concentration at 0 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of Ova displayed as the title of each subplot.

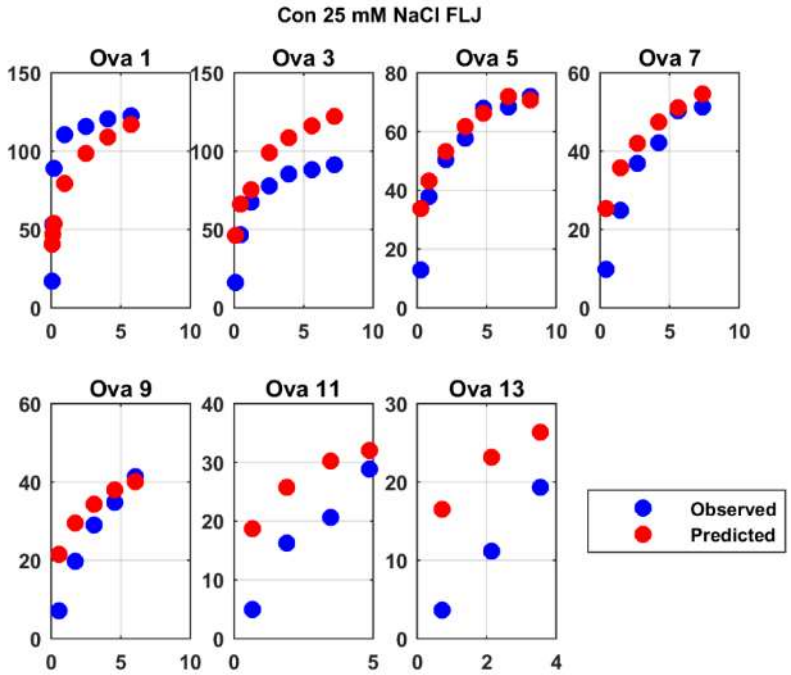

Figure 11-62 Freundlich-Langmuir-Jovanovic (FLJ) fit to binary Ova-Con binary isotherm showing observed and predicted values for Con adsorbed concentration at 25 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of Ova displayed as the title of each subplot.

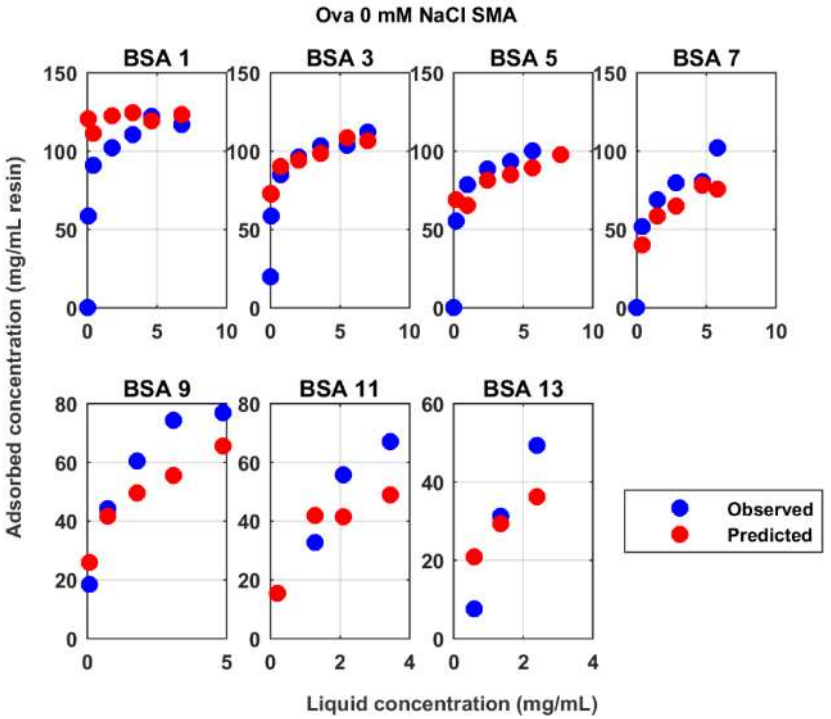

Figure 11-61 Steric mass action (SMA) fit to binary BSA-Ova binary isotherm showing observed and predicted values for Ova adsorbed concentration at 0 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of BSA displayed as the title of each subplot.

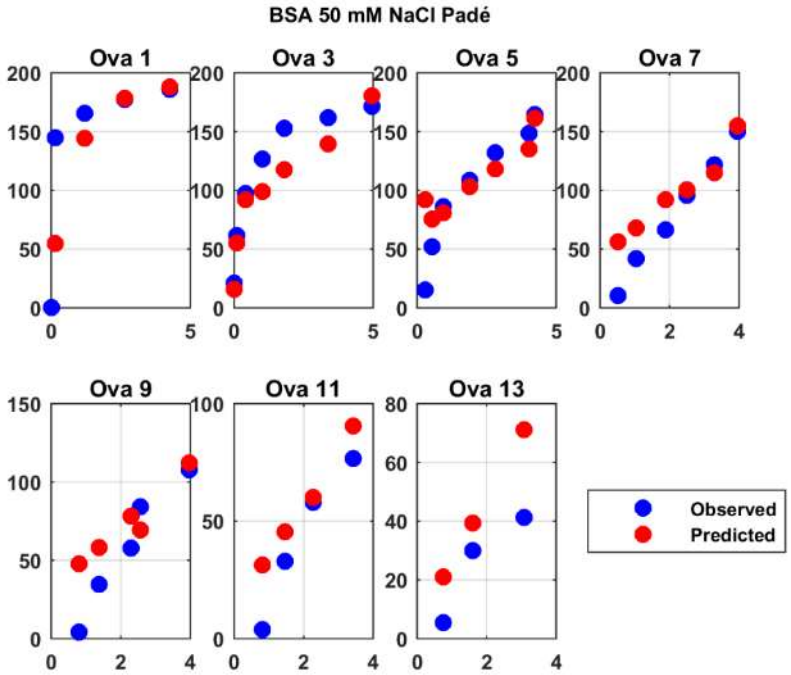

Figure 11-63 Padé fit to binary BSA-Ova binary isotherm showing observed and predicted values for BSA adsorbed concentration at 50 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of Ova displayed as the title of each subplot.

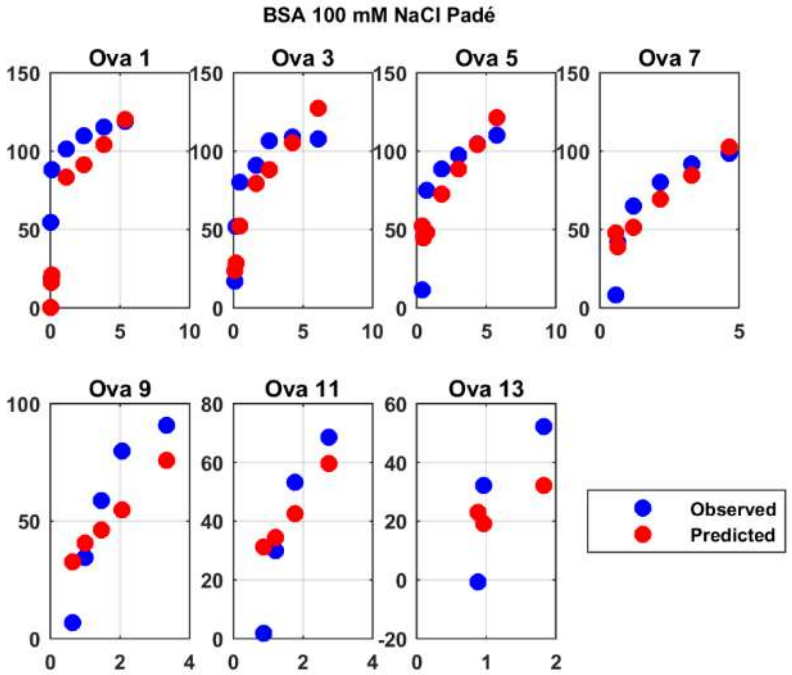

Figure 11-64 Padé fit to binary BSA-Ova binary isotherm showing observed and predicted values for BSA adsorbed concentration at 100 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of Ova displayed as the title of each subplot.

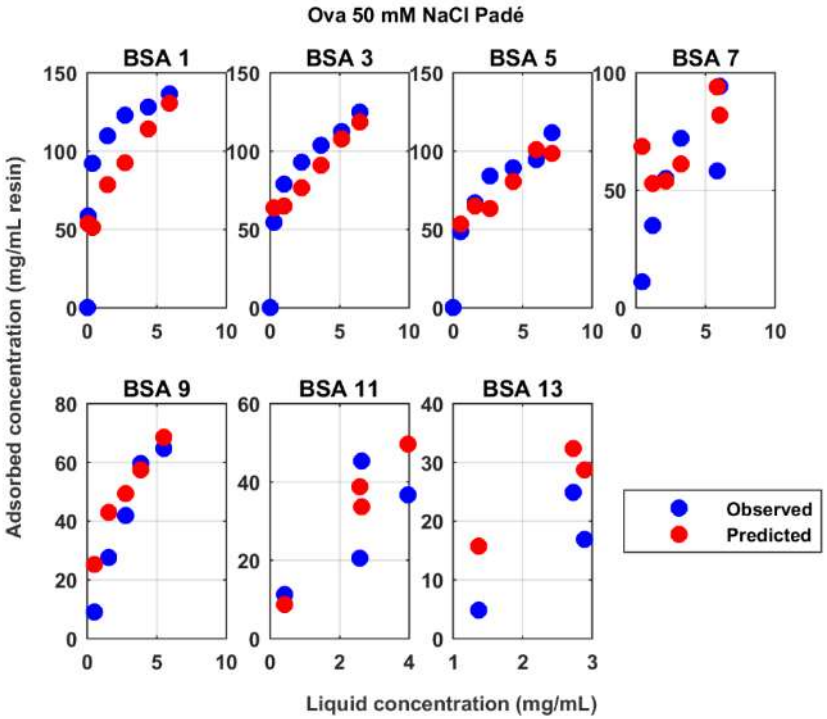

Figure 11-65 Padé fit to binary BSA-Ova binary isotherm showing observed and predicted values for Ova adsorbed concentration at 50 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of BSA displayed as the title of each subplot.

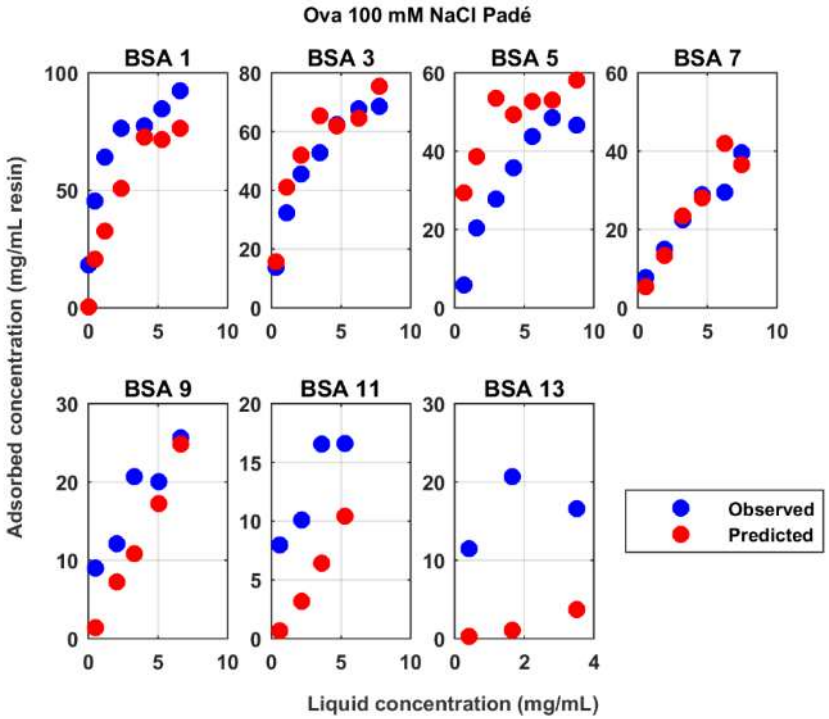

Figure 11-66 Padé fit to binary BSA-Ova binary isotherm showing observed and predicted values for Ova adsorbed concentration at 100 mM NaCl. The binary isotherm has been broken down into a series of single component isotherms with each subplot showing an increasing starting concentration of BSA displayed as the title of each subplot.

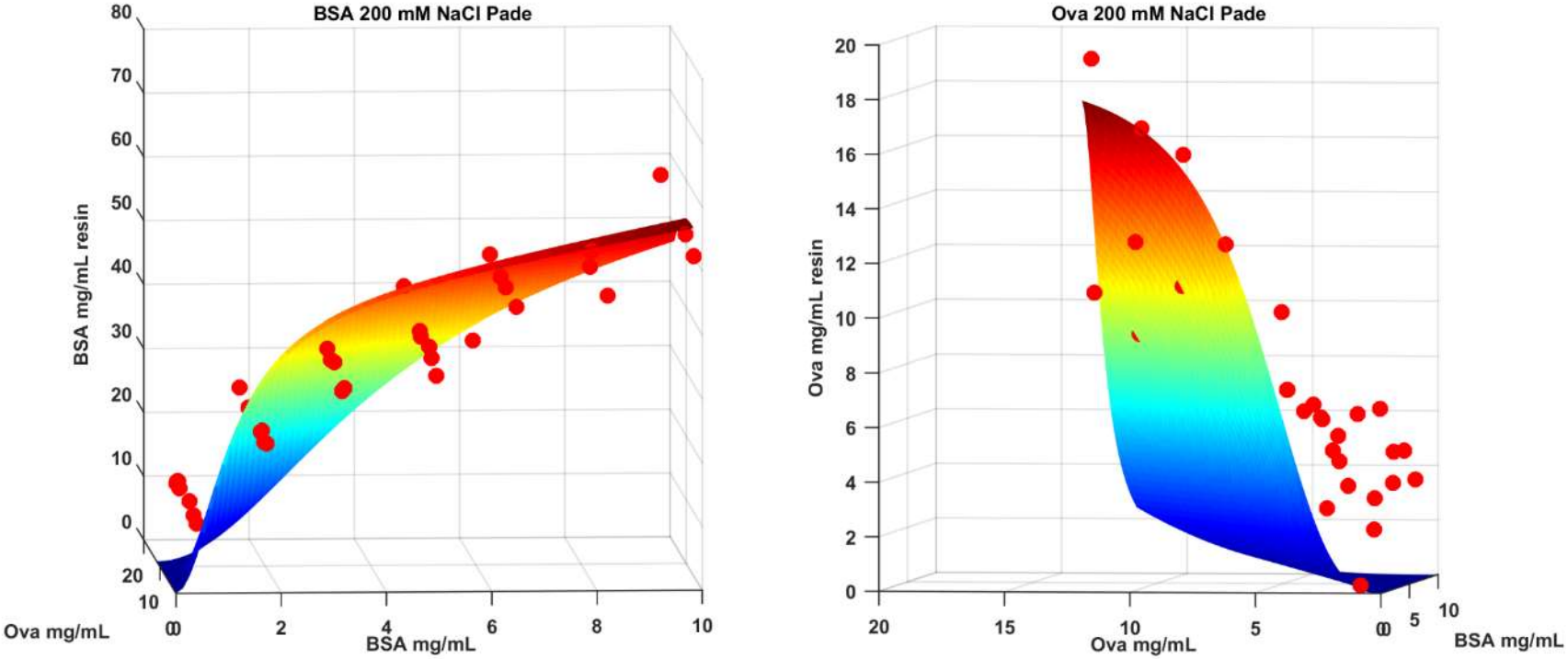

Figure 11-67 BSA-Ova binary isotherm displaying the Padé surface fit at 200 mM NaCl. Left hand panel displays BSA adsorbed concentration and right hand panel adsorbed Ova. Experimental data has been ploted as red points.

12.0 Appendix D Publications from this thesis